Quantifying the Life Stages of a Biomolecule: Implications for the Circadian Transcriptome

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. rer. nat.
im Fach Biophysik

eingereicht an der
Lebenswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
Dipl. Phys. Sarah Lück

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Bernhard Grimm

Gutachter:
1. Prof. Dr. Hanspeter Herzel
2. Dr. Pål O. Westermark
3. Prof. Dr. Felix Naef

eingereicht am: 11.01.2017
Tag der mündlichen Prüfung: 02.06.2017

Abstract

In almost all organisms on Earth, many behavioral, physiological, and biochemical activities oscillate with a circadian rhythm, a rhythm with a period of about 24 hours. This oscillating behavior is the result from a combined influence of the external day-night-cycle and an internal timekeeping system.

In gene expression, the 24-hour-rhythm can be found on all stages: from transcription initiation to protein degradation. On the transcript level, circadian mRNA production and mRNA abundance are comprehensively charted through numerous genome-wide high throughput studies from various model organisms and tissues. Circadian post-transcriptional regulation, however, comprising many different processes, is less well understood. In this thesis, I will investigate how unobserved post-transcriptional processes influence rhythmic properties of gene expression. To this end, I quantify the life-stages of biomolecules using one modeling motif, a simple ordinary differential equation describing production and degradation with time-dependent rhythmic rates. This basic modeling motif is systematically varied to examine and discuss various influences of post-transcriptional regulation (PTR) on circadian mRNA expression.

First, the influence of PTR on the circadian transcriptome is theoretically investigated, specifically, the effects on phase and amplitude of transcript abundance. Constant PTR dictates restrictions on phase and amplitude relations between transcript production and abundance: mRNA can peak at most 6 hours (in circadian context) after its production, the mRNA amplitude is smaller than the production amplitude. However, genome-wide studies of production and abundance show many genes where these relations do not hold true, and thus an oscillation also in a post-transcriptional process must be assumed. A careful discussion of known post-transcriptional processes suggests that only rhythmic mRNA degradation, rhythmic mRNA export or rhythmic alternative splicing are able to explain measured circadian expression profiles.

The model provides the basis for a statistical test to quantify the extent of rhythmic PTR in genome-wide studies. Analyzing two data sets on mouse liver and kidney, I find that 18% of circadian genes in kidney and 34% in liver are under rhythmic post-transcriptional control. The untranslated regulatory regions (UTR) of circadian mRNAs are longer in liver compared to kidney and phase analysis points to a peak in mRNA degradation around CT12 for liver and around CT0 in kidney.

In a second part, I analyze more specific aspects of PTR in a hypothesis-driven approach. Firstly, I find that splicing with a rhythm of 24 hours is able to generate 12-hour rhythms in abundance of mature mRNA, and I further characterize the requirements for this phenomenon. Secondly, I propose and analyze a model to investigate rhythmic degradation of core clock genes. And finally, I extend the core modeling motif to a partial differential equation (PDE) model that accounts for the "aging" process of molecules. I first use the PDE to investigate oscillations in oxidized proteins: Long-lived proteins tend to "forget" their production history and – although rhythmically produced – proteins have a constant total abundance. However, the production rhythm is still observable in fractions of damaged proteins and this might have biological implications. As a second application of the PDE, I describe the time course of poly(A) tail length distributions. The Poly(A) tails is a stabilizing element of mRNA. This description is motivated by novel methods to measure poly(A) tails genome-wide (e.g. "TAIL-seq") and can be used to predict poly(A) tail deadenylation rates from such data. Here, the original modeling motif finds a new and independent application apart from describing circadian gene expression.

In this thesis, I varied a minimal modeling motif to query large datasets for evidence of specific hypotheses on underlying mechanisms. This is essentially an Occam's razor approach: only very minimal assumptions are made and rigorously tested. The approach offers a promising general solution, unifying the disparate scientific approaches of data-driven vs. hypothesis-driven research in molecular biology.

Zusammenfassung

Viele biologische Prozesse im Verhalten von ganzen Organismen, aber auch in den Prozessen und der biochemischen Zusammensetzung von Zellen zeigen einen zirkadianen Rhythmus, also einen Rhythmus mit einer Periode von etwa 24 Stunden. Diese Oszillationen sind das Resultat vom äußeren Tag-Nacht-Rhythmus und einer inneren biologischen Uhr.

Diese 24-Stunden-Rhythmen sind in der Genexpression auf allen Ebenen zu finden: von der Transkriptionsinitiation bis zur Proteindegradation. Auf Transkriptebene, zirkadiane mRNA-Produktion und mRNA-Abundanz ist umfassend durch zahlreiche genomweite Hochdurchsatzstudien in mehreren Modellorganismen und Geweben gemessen. Auf der anderen Seite, zirkadiane posttranskriptionelle Regulation, die verschiedenste Prozesse umfasst, ist weit weniger verstanden. In dieser Arbeit untersuche ich, wie bisher ungemessene, posttranskriptionelle Prozesse die rhythmischen Eigenschaften von Genexpression beeinflussen. Dazu beschreibe ich die Lebensstadien eines beliebigen Bio-Moleküls mit einem Modell-Motiv, einer einfachen Differentialgleichung mit zeitabhängigen, rhythmischen Raten. Diese Differentialgleichung wird variiert um systematisch den Einfluss von posttranskriptioneller Regulation (PTR) auf zirkadiane Genexpression zu untersuchen.

Als erstes untersuche ich theoretisch den Einfluss von PTR auf das zirkadiane Transkriptom, speziell den Effekt auf Phase und Amplitude. Konstante PTR setzt Einschränkungen für die Phasen- und Amplitudenbeziehung zwischen mRNA-Produktion und mRNA-Abundanz fest: mRNA kann höchstens 6 Stunden (im Kontext der zirkadianen Uhr) nach der Produktion ihren Hochpunkt erreichen, die mRNA-Amplitude ist kleiner als die Produktionsamplitude. Genomweite Studien zeigen jedoch, dass für viele Gene diese Beschränkungen nicht erfüllt sind. Diese Eregbnisse können nur erklärt werden, wenn es auch eine Oszillation in einem posttranskriptionellen Prozess gibt. Eine sorgfältige Diskussion bekannter posttranskriptioneller Prozesse zeigt, dass nur rhythmischer mRNA-Abbau, rhythmischer mRNA-Export und alternatives Spleißen die gemessenen zirkadianen Expressionsprofile erklären können.

Das Modell liefert die Grundlage für einen statistischen Test um das Ausmaß von rhythmischer PTR in genomweiten Studien zu quantifizieren. Durch die Analyse zweier Datensätze von Mausleber und -niere finde ich, dass 18% aller zirkadianen Gene in Niere und 34% in Leber rhythmisch posttranskriptionell reguliert sind.

Im zweiten Teil analysiere ich weitere Aspekte von PTR in einem Hypothesen-getriebenen Ansatz. Als erstes weise ich nach, dass Spleißen mit einem Rhythmus von 24 Stunden 12 Stunden-Rhythmen in der Abundanz von reifer mRNA erzeugen kann, das heißt die Abundanz hat zweimal pro Tag einen Hochpunkt. Zusätzlich charakterisiere ich die Bedingungen für dieses Phänomen. Als nächstes schlage ich ein Modell vor, das rhythmische Degradation von Mitgliedern der zentralen inneren Uhr, der zirkadianen Uhr, beschreibt, und charakterisiere dieses. Schließlich erweitere ich das Modell-Grundmotiv zu einer partiellen Differentialgleichung (PDG), die das “Altern” von Molekülen beschreibt. Ich schlage zwei Beispiele für die Anwendung vor. Als erstes benutze ich die PDG um Oszillationen in oxidierten Proteinen zu untersuchen: Langlebige Proteine “vergessen” ihre Produktionsgeschichte und weisen trotz rhythmischer Produktion ein konstantes totales Proteinlevel auf. In Anteilen von beschädigten Proteinen können aber die Produktionsrhythmen noch beobachtet werden, was biologische Auswirkungen haben kann. Als zweites Beispiel beschreibe ich die Längenentwicklung von Poly(A)-Schwänzen, ein stabilisierendes Element von mRNA-Molekülen. Diese Beschreibung ist motiviert durch eine neue Messmethode, die genomweit die Längenverteilung von Poly(A)-Schwänzen misst (“TAIL-seq”), und dazu benutzt werden kann, von solchen Daten die Poly(A)-Schwanz-Deadenylierungsrate vorherzusagen.

In dieser Arbeit variiere ich ein minimales Modellierungsmotiv um große Datensätze nach Hinweisen von darunterliegenden Mechanismen zu untersuchen. Damit folge ich im Wesentlichen dem Ockhams-Rasiermesser-Prinzip: Es werden nur minimale Annahmen gemacht und gründlich getestet. Dieser Ansatz bietet eine vielversprechende, generelle Lösung, die die ungleichen wissenschaftlichen Ansätze von datengetriebener vs. hypothesengetriebener Forschung in der Molekularbiologie vereinigt.

Abandon the urge to simplify everything, to look for formulas and easy answers, and to begin to think multidimensionally, to glory in the mystery and paradoxes of life, not to be dismayed by the multitude of causes and consequences that are inherent in each experience – to appreciate the fact that life is complex.

M. Scott Peck

Everything must be made as simple as possible. But not simpler.

Albert Einstein

The scatter of the experimental data contrasted with the convincing clarity of the theoretical model.

Matthias Beuse *et al.* [1]

Contents

1 Introduction

1.1 Biological 24-Hour Rhythms

For more than three billion years, the organisms on this planet have known, just like Little Orphan Annie, that "The sun'll come out tomorrow", and many have honed their biochemistry to exploit this knowledge.[1] At this time *Cyanobacteria*, one of the oldest clades on Earth, started to use water as the electron-donor for photosynthesis, releasing oxygen, and thus slowly building up the earth's atmosphere. About 1 billion years ago these bacteria developed an internal timekeeper. With this internal "clock" they were able to synchronize the timing of their metabolic events with the predictable turning of the earth and its implications for daily changes in light, temperature and humidity. Such clocks are nearly ubiquitous among existing higher organisms, plants, fungi and animals [3].

Such inner clocks regulate a diverse range of cellular and organismic processes, the so-called "circadian behavior". For a process to be considered as circadian it must possess three characteristic properties [4]:

- It oscillates with a period of around 24 hours ("circa dies" - about a day) even in the absence of environmental cycles.
- The phase can change or "be reset" upon environmental cues such as light or temperature. The behavior is "entrainable".
- The period does not change for different temperatures, *i.e.* it does not run faster in higher temperatures or slow at lower temperatures. The behavior is "temperature-compensated".

In mammals, the internal time keeping system is complex and comprises of hierarchical structure, reviewed in Section 1.1.4. On the cellular level, this hierarchical system results in 24-hour-oscillations in the abundance of thousands of transcripts, see also Section 1.1.5. Genome-wide studies also found oscillating abundances in nascent RNA [5, 6]. However, these findings suggest that many of the rhythmic RNA levels cannot originate solely from their production. Many RNAs were found to oscillate on the level of mature transcript concentration, although they were not rhythmically produced, in other cases production oscillated with a smaller amplitude than the final transcript abundance; or, in some case both, production rate and transcript abundance oscillated, but the phase difference between production and transcript abundance was greater than 6 hours. In all these cases, post-transcriptional processes must play an important role

[1] I stole this sentence from Susan Golden's review on Cyanobacterial circadian clocks [2]. The song comes from a musical "Annie" composed by Charles Strouse and lyrics by Martin Charnin. The musical is based on a comic strip series which was published from the 1920s until mid 1970s. It was created by Harold Gray.

in shaping the circadian abundance profile of the biomolecule. These results were the starting point for our investigation of rhythmic post-transcriptional regulation in more depth. The weapon of choice when tackling the diverse scientific questions that arise was mathematical modeling. Specifically, I described rhythmic processes with ordinary differential equations or partial differential equations. The results of this journey are concentrated in this doctorial thesis.

Before we start our exploration of the varied aspects of rhythmic post-transcriptional regulation, I will first give an introduction to the circadian clock, in particular the mammalian clock, review what current knowledge of rhythmic post-transcriptional regulation gained from experimental studies. I will then introduce the central theoretical concepts, such as amplitude and phase, which will be employed throughout this thesis.

1.1.1 A Brief History of Chronobiology

The existence of internal time keepers was originally proposed in the year 1729. De Mairan [7] noticed that the plant *Mimosa pudica* lowers its leaves and folds them away during night. He also found that this behavior continues in constant darkness. This opened the question on the existence of internal timekeepers. Almost 200 years later, the search for clocks started to spread. Several studies were undertaken (in chronical order) into: primates [8] (1906), rodents [9, 10, 11], insects [12], birds [13], single-cell eukaryotes [14], 1962 by Aschoff *et al.* [15] humans, fungi [16] and finally, 1986, also bacteria [17]. All these studies tested various behaviors which continued to oscillate with a period of around 24 hours in darkness or in different light environments, strengthening the inital idea of the existence of internal time-keepers in all domains of life. 1959, Franz Halberg introduced the term "circadian" derived from the two Latin words "circa" and "dies" accounting for oscillations with a period of "about a day".

Evidence for a genetic base of the internal time keeper was presented in 1935 by Bünning [18] who showed that period length is inheritable in bean plants. A genetic component, however, was first identified in the late 20th century. Here, Konopka *et al.* [19] identified a gene locus in *Drosophila Melanogaster* where mutations resulted in either longer or shorter periods of pupal eclosion and locomotor activity. The encoded gene *Per* (*Period*) was characterized 10 years later [20]. In mice, a large mutant screen in the laboratory of J. Takahashi lead to the discovery of the first clock gene, named *Circadian Locomotor Output Cycles Kaput* (*Clock*) in the mid 90s [21]. Only 5 years later, the main components of the clocks in mammals, *Drosophila melanogaster*, *Neurospora crassa* and *Synechococcus* were characterized [3].[2]

1.1.2 "Evolutionary Why" Clocks?

The first circadian clocks are thought to have evolved 1 billion years ago [23] in cyanobacteria, one of the oldest clades on earth. In these bacteria, cycles in the phosphorylation state of the protein KaiC drive alternation between periods of nitrogen fixation and photosynthesis [2]. Additionally, a circadian cycle of genome compaction, potentially driven by changes in the extent of DNA supercoiling, has been observed in a variety of species [24, 25]. In the early days of Earth

[2] Of course, these two paragraphs are a very condensed summary and can only provide a glimpse on the development of research of the circadian clock. For further reading, I highly recommend the comprehensive and entertaining introduction to chrono-history in the book "Circadian Physiology" by Roberto Refinetti [22].

ultra-violet (UV) irradiation from the sun - not yet filtered by the Earth's atmosphere and its ozone layer - was both a threat and the main energy source. One theory for the emergence of clocks in cyanobacteria is the "escape from light" hypothesis. This hypothesis states that clocks evolved to temporally separate cellular processes that are sensitive against UV radiation, such as DNA replication, from other, UV-insensitive processes. Hence, the DNA may be compacted during daytime in order to protect the genomic information. Records of DNA compaction and decompaction, however, show that, in modern cyanobacteria, compaction occurs at night, not during the day [26]. Hut *et al.* [27] hence argue that clocks have evolved in order to store ATP for periods where no photosynthesis is possible. To date it is unclear which of the two theories holds true.

Clocks in plants, fungi, and animals are quite distinct from each other and have no apparent homologies with KaiC from cyanobacteria [3, 28, 29]. This indicates that clocks have evolved more than once. It is interesting to note that cryptochrome, a light-sensitive protein and main component of the circadian clock in mammals [30, 31] and *D. Melanogaster* [32, 33], has a close relationship to photolyases, an ancient enzyme family that uses light to repair UV-induced DNA damage [34]. Further, it has been observed that despite their different clocks in all domains of life oxidation-reduction cycles of peroxiredoxin proteins oscillate with a circadian period [35]. Peroxiredoxin proteins protect cells from excessive amounts of reactive oxygen species (ROS), which are produced during respiration and photosynthesis. These observations fit with the "escape from light" hypothesis and may suggest a primordial role of DNA damage in the evolution of circadian clocks in eukaryotes [36, 37].

Other lessons may be learned from another exisiting clock. Budding yeast has evolved a metabolic cycle which is tightly coupled to the cell division cycle [38, 39, 40, 41]. The cycle shares a variety of features with the circadian clock, discussed in Causton *et al.* [42], and both clocks might even share a common ancestor [43, 44]. The metabolic oscillations were first observed in a synchronous oscillation in the pH value and dissolved oxygen levels in continuously growing cultures of budding yeast [45, 38], but could recently be observed in single cells growing in a microfluidic system [41]. Akin to the "escape from light" hypothesis, it was proposed that this metabolic cycle may serve to protect DNA from ROS during replication [39]. However, oscillations were also observed in non-growing cells [46], and it was recently shown that phases of high oxygen consumption, where ROS are produced, are not strictly separated from DNA replication [40, 47]. These results put forward original interpretations of this cycle from the 1960s [38, 48], that initiation of cell division (in budding yeast: "budding") and DNA replication occur only after accumulation of sufficient carbohydrate reserves to supply the increased energy demand during these time-critical cellular processes. Machne *et al.* [49] proposed the existence of a global mechanism of growth regulation, where a direct feedback exists between energy metabolism, ATP-dependent remodeling of chromatin structure, and large groups of differentially expressed genes in budding yeast. They further suggested a similar system to be tightly integrated with the circadian clock of cyanobacteria, where DNA structure ("DNA supercoiling" and DNA compaction) were previously observed to oscillate in a circadian manner.

However, energy metabolism is a major source of ROS. With this in mind, it can be argued, that protection from ROS- or UV-induced DNA damage, or gating replication for times of high energy supply may be two sides of the same came coin and hence both "evolutionary whys" are

compatible.

1.1.3 Architecture of Cellular Clocks

Circadian clocks in multi-cellular organisms consists of several intertwined transcriptional-translational feedback-loops. A genome-wide small interfering RNA (siRNA) screen showed that more than 200 genes affect properties of mammalian circadian gene expression, such as period and amplitude of mRNA abundance [50]. Hence, a clock network is large and complex becoming easily incomprehensible. The consensus view focuses therefore on small sets of a few transcriptional (co-)regulators, so-called core clock proteins, that operate in most cell types and necessary for a functioning cellular clock.

In plants, fungi, and animals these core-clock genes act as transcriptional activators and repressors, mutually regulating each other in negative transcriptional-translational feedback loops, as has been shown in several key experiments [51, 52, 53]. A negative feedback loop is one major ingredient to produce oscillations, also comprehensively understood from the modeling perspective [54, 55, 56]. The other two ingredients for oscillation production are delay and a sufficient non-linearity in the describing equations. A delay results from the inclusion of intermediate steps within the negative feedback loop. While a single transcriptional auto-repressor acts too quickly on its own transcription, a chain of cross-regulating transcription factors can easily cause oscillatory expression patterns [57, 58]. Non-linearity is introduced by cooperativity in transcriptional activation or suppression, *i.e.* more than one protein at a time is required to activate or inhibit transcription. Because of this, the *e.g.* inhibition of transcription, which is dependent on the concentration of inhibitory proteins, follows a step function, which is highly nonlinear. Delay and non-linearity are inversly dependent on each other. If the delay is large, less non-linearity is needed to produce oscillations [56]. In many cases, circadian clocks are equipped with additional positive feedback loops. These might weaken the requirements on nonlinearity and provide robustness with regards to degradation parameters [59].

In mammals, the proteins BMAL1 and CLOCK form a dimer activating the transcription of - among others - the transcriptional co-repressors PER1/2 and CRY1/2 which in turn inhibit the activation of BMAL1/CLOCK and thereby exert a negative feedback on their own expression, see Figure 1.1. The consensus view considers this to be the main negative feedback loop in the mammalian circadian clock [60, 28] supported by the fact that a knock-out of either one of its components, namely BMAL1 [61], CLOCK and its substitute NPAS2 [62], PER1 and PER2 [63, 64] or CRY1 and CRY2 [65, 66], completely stops the clock. However, a very recent computational approach pointed to a different main negative feedback-loop. Here, Pett *et al.* [67] systemically substituted oscillations of core-clock members by their mean mRNA concentration in a comprehensive, already published mathematical model of the circadian clock [68] and investigated if circadian-like oscillations still persist. With this approach Pett *et al.* [67] were able to reduce the circadian network structure to essential parts. A motif appeared in this systematic analysis was the “repressilator”, a network motif consisting of a serial inhibition [69], and originating in Goodwin’s model of cellular oscillators [57]. The identified repressilator motif Pett *et al.* [67] represented the serial inhibition of the clock genes $Cry1 \rightarrow RevErb\alpha \rightarrow Per2 \rightarrow Cry1$, see Figure 1.1. It is interesting to see, that PER2 and CRY1 in the model described previously form a dimer acting as one transcriptional co-repressor, but their DNA binding profiles

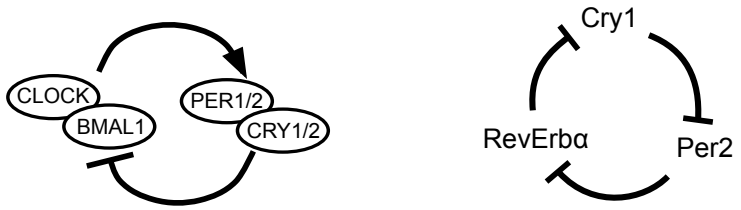


Figure 1.1: **Two models of the core negative feedback loop in the circadian clock.** The consensus view favors the idea shown in the left. The proteins BMAL1 and CLOCK form a dimer which activates the transcription of the transcriptional co-repressors PER1/2 and CRY1/2 which in turn inhibit the activation of BMAL1/CLOCK. Recently, Pett *et al.* [67] identified a repressilator (right) as core negative feedback of the circadian clock.

tend to support sequential DNA binding [70] of both proteins. These binding profiles favor the repressilator model over the consensus feedback-loop. Knock-out studies further support the repressilator motif as the core negative feedback loop, see [67] and references therein. However, the model and hence finding a repressilator as a possible main negative feedback-loop is based on mRNA expression profiles and does not take protein levels into account. Including protein levels might, yet again, alter the picture.

In summary, the overall architecture of the mammalian core clock is well understood and the details are currently worked out in an exemplary loop between theoretical modeling and wet lab experimentation.

1.1.4 Time-Keeping in Higher Organisms

Every cell with a nucleus is equipped with the core clock described above. However, even cells without nucleus, *i.e.* blood cells show sustained 24-hour-oscillations in their peroxiredoxin protein levels [71] indicating the existence of additional oscillators within single cells [72]. Hence, most, if not all, cells show oscillations, even in the absence of transcription. To function as a proper time-keeper these oscillators must transmit their information about time to other cellular processes and they must listen to their environment in order to synchronize to the external cues ("entrainment"). How is this achieved?

In mammals - highly complex organisms - single cell oscillators are organized in a hierarchical structure. Single cells group together to form larger structures. It is believed that these larger structures divide in so-called "peripheral oscillators", associated with whole organs such as liver or kidney, and "master oscillators", which respond to environmental cues. The peripheral oscillators listen to the master oscillators.

A master oscillator, or central pacemaker, is able to restore complex rhythmic behavior in modified, arrhythmic animals and is connected to the environment [73]. There are three knwon master oscillators, the suprachiasmatic nuclei (SCN), the food entrainable oscillator (FEO) and the methamphetamine sensitive circadian oscillator (MASCO).

From these three, the SCN is best understood. It is a small organ in the hypothalamus and consists of about 20,000 neurons [74]. The individual clocks of these neurons are tightly

coupled such that each neuron in the SCN oscillates with almost exactly the same period [75] and therefore with a fixed phase relationship. The SCN receives information directly via light reaching the retina of the eye [76]. The SCN reacts upon different photoperiods, accounting for long or short days, with modulated phase distributions of its neurons [77, 78], also which has also been investigated from a modeling perspective [79]. The SCN transmits its timing information via neuronal and humoral signals [76], which are yet incompletely understood [73]. SCN lesioned animals have arrhythmic locomotor activity [80]. A transplant of an intact SCN restores rhythmicity [81], with the original period of the transplanted SCN [82]. All this together, the coupling to the environment and the capability to induce rhythms in complex behavior, support the role of the SCN as a pacemaker and rhythm provider.

The two other pacemakers, FEO and MASCO, investigated and put forward by Michael Menaker and colleagues, are less well understood. Evidence for their existence is based on experiments where rhythms in locomotor activity, body temperature and gene expression profiles can be restored in SCN-lesioned animals by restricted feeding [83, 84, 85, 86] or methamphetamine exposure [87, 85, 86]. In SCN-intact animals a phase de-synchrony between food-dependent and food-independent tissues is induced by restricted feeding [88]. In genetically modified mice lacking a functional transcriptional-translational clock, rhythmic locomotor-activity is induced both by methamphetamine administration [89] and restricted feeding [90], suggesting that the molecular mechanisms that generate these oscillations are fundamentally different from those that generate SCN rhythmicity. However, neither the anatomical location nor the molecular mechanisms of the FEO or the MASCO are yet known. They might even be two aspects of the same mechanism [73]. However, these experiments strongly support the idea that oscillators other than the SCN can act as pacemakers.

Peripheral oscillators on the other hand are rhythm listeners. Individual organs are thought to have their own clock which coordinates tissue-specific behavior and which is induced by cues from the SCN and other pacemakers [91, 92, 93]. Every organ shows its own specific circadian transcriptome profile [94] accounting for different organ-specific functions, see also Sec. 1.1.5. If peripheral organs, *i.e.* organs beside the SCN, are explanted they show damped oscillations [95]. This is in contrast to SCN explants which were reported to oscillate indefinitely [95]. Theoretically, the observed dampening could either arise from attenuated oscillations in single cells or from increasing de-synchrony among clocks in adjacent cells. Bio-luminescent or fluorescent recordings of fibroblasts cells in culture [96, 97] and of different tissue explants [98] supports the latter hypothesis: Single cells show long-lasting oscillations - each with a slightly different period - but fail to synchronize without external stimuli. Westermark *et al.* [99] compared time series of SCN neurons and fibroblasts to two models, firstly a damped oscillator driven by noise, secondly a self-sustained noisy oscillator. These two models account for a "weak" and a "strong" oscillator, comparable to the concept of peripheral (weak) oscillator and (strong) pacemaker. However, it was found that single cell time series from fibroblasts and SCN neurons were both described by the two oscillators, indicating that both cell types contain similar clocks. Since SCN and peripheral organs are similar at the single-cell level, the differing behavior of tissue explants, non-dampened *versus* dampened rhythms, most likely arises from a missing intercellular coupling in peripheral organs. Consequently, there is no experimental evidence that peripheral organs form one meta-oscillator. They rather consist of single-cell oscillators which are inde-

pendently entrained by the pacemakers. Different cell types in heterogeneous tissues might be entrained by different input-pathways. These results would suggest that the concept of one clock per organ, *e.g.* "liver clock", may be misleading.

Furthermore, there is evidence for a secondary hierarchical structure within peripheral organs. For example, the adrenal gland which "listens" to the SCN [100, 101], produces hormones which in turn influence rhythmicity in other organs such as liver or kidney [102, 103, 104, 105]. It may even be possible that several more layers exist within this hierarchy.

Consequently, an oscillating transcriptome measured in a peripheral tissue reflects the influence of a myriad of inputs from different levels of organization: from the transcriptional-translational negative feedback loop, the core clock, to hormonal and neuronal inputs of the pacemakers, and environmental cues such as metabolic activity upon feeding or locomotor activity. The transcriptome of cell lines, a system where inputs from pacemakers do not exist, showed rhythms in about only a dozen transcripts, in liver tissue however rhythms in over 3000 transcripts could be detected [106]. This indicates the large majority of circadianly expressed genes in tissues is directly influenced by systemic cues from the pacemakers.

1.1.5 Circadian Output

Depending on tissue, experiment and analysis (see next section) 5% to 50% of transcripts show a circadian rhythm in their abundance. Depending on protein half-lives and cell growth, many cellular processes could thus be modulated by a circadian rhythm. This includes both basic cellular processes, which are needed for a functioning cell, and functions specific for certain cell types and organs. Additionally, as outlined above, transcription-independent oscillations further point to autonomous metabolic or physiological clock-like systems. However, very little is known about the mechanics and main actors in these systems.

Basic cellular functions are under circadian control Cellular metabolism and the core circadian clock are interlocked. Most of cellular metabolism oscillates with a circadian rhythm and it is influenced by members of the core clock. This includes glycolysis [107], gluconeogenesis [108, 109, 110], fatty acid oxidation [111] and lipid biogenesis [112]. Thus, both build-up and consumption - anabolism and catabolism - of the main cellular energy sources are controlled by the clock. It has also been shown that mitochondrial respiration activity [113, 114] is circadianly regulated, both by controlling rate-limiting enzymes [115, 114] and mitochondrial dynamics [113], *i.e.* fusion and fission of the mitochondrial network, which is connected to the mitochondrial activity [116]. Consequently, ATP levels are circadian in a multitude of tissues [117]. The metabolic state of a cell feeds back to the core clock via the NAD+-NADH axis, reviewed in [118].

Cellular maintenance processes, often connected to the bioenergetic state of the cell [119, 120, 121], are under circadian control, such as autophagy [122] and the cellular redox state, exquisitely reviewed by Putker *et al.* [120]. Thus, probably the whole bioenergetic state of most - if not all - cells oscillates with a diurnal rhythm, reminiscent again of the metabolic oscillations in budding yeast cells outlined above. And just like in these metabolic oscillations, the other large oscillator in the cell, the cell cycle, and the circadian clock are coupled [123, 124].

But cells do not simply exist, they must also fulfill a function in a complex organism. How does the clock affect tissue specific functions, especially in liver and kidney, the two organs I analyze in this thesis with regard to their post-transcriptional regulation? According to Zhang *et al.* [94] liver and kidney are the two organs with the highest fraction of circadianly expressed genes, *i.e.* 16% and 13% of expressed genes in liver and kidney, respectively, have a circadian rhythm in their transcript abundance. What does this mean for organ function?

Liver - virtually all processes oscillate The liver is the second largest organ in our body after the skin and mainly consists of hepatocytes. It is a metabolic organ that performs numerous functions in the human body, including regulation of glycogen storage, decomposition of red blood cells, plasma protein synthesis, hormone production, and detoxification. The liver is considered the prototypic experimental model tissue for circadian research in peripheral organs. Hence, numerous studies have covered the circadian transcriptome [125, 126, 127, 128, 129, 130, 131, 5, 6, 132, 133, 94], proteome [134, 135, 136, 137] and metabolome [130, 138, 139, 140, 141] of the liver. Many key findings for the global regulation of circadian gene expression were generated using mouse liver as modelling system [6, 5, 123, 135]. But the data also served for elucidating circadian hepatic processes. Virtually all hepatic processes appear to be under control of the circadian clock, including detoxification, cholesterol and bile acid metabolism, glucose and lipid homeostasis, thyroid hormone metabolism, and many more (reviewed in [142]).

Kidney function is modulated with diurnal rhythm The kidneys are two bean-shaped organs, which function to filter water soluble waste products of metabolism from the blood, maintain fluid and ion homeostasis and regulate blood pressure. In contrast to the liver, a kidney is a very heterogeneous organ and consists of many different cell types. The functional filtering unit in kidney is the tubular structured nephron. Each kidney contains about 1 million nephrons. Blood enters the nephrons and is filtered through the glomerulus, a ball-shaped structure within the nephron. The filtrate flows through the nephron, and the many specialized cell types in each segment of the nephron reabsorb or secrete solutes according to the needs of the body. The final filtrate flows into the ureter to eventually become the urine.

The overall kidney function seems to exhibit circadian fluctuations. For example, the glomerular filtration rate, *i.e.* the flow rate of filtered fluid through the kidney [143], and the renal blood flow [144] show diurnal oscillations.

Urinary sodium excretion oscillates with a diurnal rhythm [145]. This might be explained by clock-mediated regulation of several renal sodium transport genes [146, 147, 148]. Several key pathways in the kidney are affected by core clock knock-outs (reviewed in [149]).

1.2 Rhythmic Post-Transciptional Regulation

Let's start this introduction by investigating the statement below:

> *Messenger RNAs (mRNAs) are the central conduits in the flow of information from DNA to protein. In eukaryotes, mRNAs are first synthesized in the nucleus as pre-mRNAs that are subject to 5'-end capping, splicing, 3'-end cleavage, and polyadenylation. Once pre-mRNA processing is complete, mature mRNAs are exported to the cytoplasm, where they serve as the blueprints for protein synthesis by ribosomes and then are degraded.*

This text snippet, taken from the excellently written review by Melissa Moore [150][3], contains most of the key words which one needs to discuss about post-transcriptional regulation and serves its purpose of introducing the topic. However, in this review it is given as a counter example which is not capable to reflect, in reality, vast and fascinating complexity of post-transcriptional regulation.

Another example for this complexity is provided by the beautiful figure of Gerstberger *et al.* [151][4] where post-transcriptional regulation of both mRNA and non-coding RNA is shown. In this paper, the authors derive a census of about 1500 RNA binding proteins (RBP). The subsequent analysis reveals that most RBP are ubiquitously expressed, typically at higher levels than average proteins, and almost all of them are involved in protein synthesis (692 mRNA binding proteins, 169 ribosomal proteins, 130 proteins in biogenesis and delivery of charged tRNAs to the ribosome).

In principle, any abundance or activity of proteins involved in post-transcriptional regulation, any depicted step in the figure, or any processing rate could exhibit circadian rhythms affecting single mRNAs, groups of mRNA or even all expressed mRNAs in specific cell types. In the following, I will briefly review what is already known regarding rhythmic post-transcriptional regulation in mammals and *Drosophila*.

Post-transcriptional regulation is affected by mRNA-associated factors, such as RNA binding proteins (RBP) and micro-RNAs (miRNAs). The mRNA-associated factors act as trans-factors that recognize cis-acting binding sites on (pre)mRNA. Rhythmic post-transcriptional regulation can be generated either by a time-dependent activity or abundance of trans-factors or by dynamic changes of cis-elements on the (pre)mRNA achieved *e.g.* by alternative splicing.

[3] Reprinted by permission from Macmillan Publishers Ltd: Science, "From Birth to Death: The Complex Lives of Eukaryotic mRNAs", Melissa J. Moore, copyright 2005.

[4] Reprinted by permission from Macmillan Publishers Ltd: Nature, "A census of human RNA-binding proteins",Stefanie Gerstberger, Markus Hafner, Thomas Tuschl, copyright 2014.

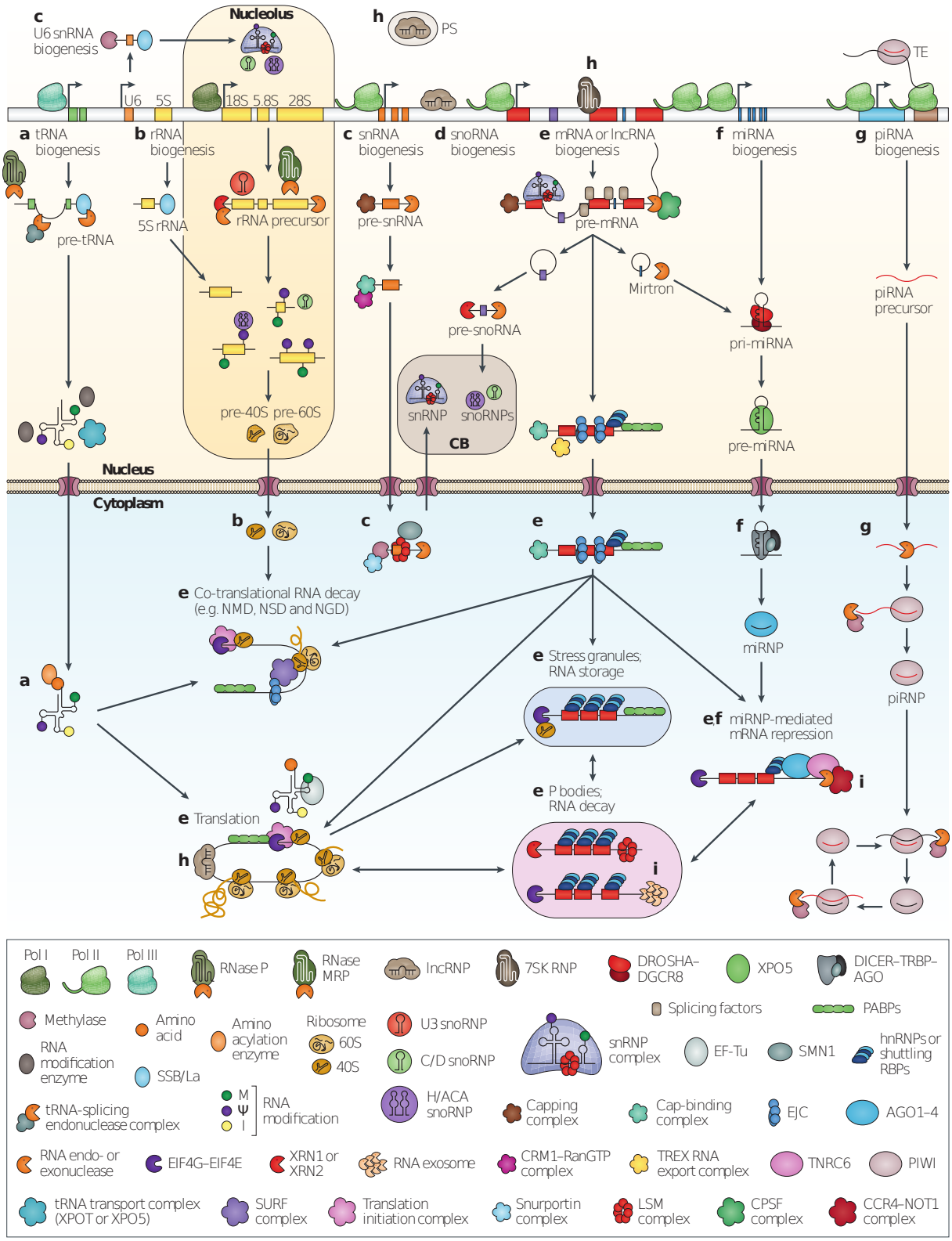


Figure 1.2: Figure caption, see next page.

Overview of the main post-transcriptional gene regulation pathways in eukaryotes. An overview is given for the biogenesis, decay and function of the most abundant RNAs: tRNAs, ribosomal RNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), mRNAs, microRNAs (miÂRNAs), PIWI-interacting RNAs (piRNAs) and long non-coding RNAs (lncRNAs). Processes are described from left to right. **a** tRNAs are transcribed by RNA polymerase III (Pol III); the 5' leader and 3' trailer sequences are removed, introns are spliced, and the ends are joined. CCA nucleotides are added to 3' ends, and nucleotide modifications - such as methylation (M), pseudouridylation (ψ) and deamination of adenosines to inosines (I) - are introduced before tRNA aminoacylation195 . **b** The 5S rRNA is transcribed by Pol III, whereas 28S, 18S and 5.8S rRNAs are transcribed as one transcript by Pol I. The precursor is processed by RNA exonucleases, endonucleases and the ribonucleoprotein (RNP) RNase MRP, guided by U3 small nucleolar RNP (snoRNP). Nucleotide modifications are introduced by snoRNPs. rRNAs are assembled together with ribosomal proteins into ribosomal precursor complexes in the nucleus and transported to the cytoplasm, where they mature to functional ribosomes. **c** Most snRNAs are transcribed by Pol II, capped and processed in the nucleus. When exported to the cytoplasm, they undergo methylation and assemble with LSM proteins into small nuclear ribonucleic particles (snRNPs) in a process aided by the survival motor neuron 1 (SMN1). These snRNPs are re-imported into the Cajal body (CB) within the nucleus, where they undergo final maturation and snRNP assembly. U6 and U6atac snRNAs are transcribed by Pol III and are alternatively processed in the nucleus and the nucleolus. Mature snRNPs form the core of the spliceosome. **d** snoRNAs and small Cajal body-specific RNAs (scaRNAs) are processed from mRNA introns, capped and modified before they assemble into snoRNPs or scaRNPs in the CB. snoRNPs and scaRNPs carry out methylation and pseudouridylation of rRNAs, snoRNAs and snRNAs, or function in rRNA processing (for example, processing of U3 snoRNA). **e** mRNAs are transcribed by Pol II, capped, spliced, edited and polyadenylated in the nucleus. Correctly matured mRNAs are exported into the cytoplasm. Regulatory RNA-binding proteins (RBPs) control correct translation, monitor stability, decay and localization, and shuttle mRNAs between actively translating ribosomes, stress granules and P bodies. **f** miRNAs are either transcribed from separate genes by Pol II as long primary miRNA (pri-miRNA) transcripts or expressed from mRNA introns (mirtrons) and processed into hairpin pre-miÂRNAs in the nucleus. After transport into the cytoplasm, they are processed into 21-nucleotide-long double-stranded RNAs. One strand is incorporated into Argonaute (AGO) proteins (forming miRNA-containing RNPs (miRNPs)) and guides them to partially complementary target mRNAs to recruit deadenylases and repress translation. **g** piRNAs are ∼28-nucleotides-long, germline-specific small RNAs. Primary piRNAs are directly processed and assembled from long, Pol II-transcribed precursor transcripts, whereas secondary piRNAs are generated in the "ping pong" cycle by the cleavage of complementary transcripts by PIWI proteins. Mature piRNAs are 2'-O-methylated and incorporated into PIWI proteins. The piRNA-PIWI complexes (piRNPs) silence transposable elements (TEs) either by endonucleolytic cleavage in the cytoplasm or through transcriptional silencing at their genomic loci in the nucleus. **h** Most lncRNAs are transcribed and processed in a similar way to mRNAs. Nuclear lncRNAs play an active part in gene regulation by directing proteins to specific gene loci, where they recruit chromatin modification complexes and induce transcriptional silencing or activation. Other non-coding RNAs (for example, 7SK RNA) reg-

ulate transcription elongation rates or induce the formation of paraspeckles (PS). Cytoplasmic non-coding RNAs can modulate mRNA translation. **i** Incorrectly processed RNAs are recognized by several complexes in the nucleus and cytoplasm that initiate and execute their degradation. CPSF, cleavage and polyadenylation specificity factor; EJC, exon junction complex; hnRNP, heterogeneous nuclear RNP; NGD, no-go decay; NMD, nonsense-mediated RNA decay; NSD, non-stop decay; PABP, poly(A)-binding protein.

1.2.1 First Advances and Investigation of Trans-Factors

The complement of possible trans-factors is large. 1500 RBPs [151] and more than 5500 mi-RNA [152] in humans have been identified, each one of them targeting groups rather than single mRNAs. It is predicted that mi-RNA affect ∼50% of all protein-coding genes [153] by downregulating gene expression via either inhibition of translation or mRNA degradation [153]. Additionally, some non-coding RNA are shown to inhibit translation [154].

The first observation of a rhythmic post-transcriptional regulation was made by Robinson *et al.* [155] who found that the mRNA encoding for SCN vasopressin had two different poly(A) tail lengths whose relative abundance varied throughout the day. They hypothesized that this is the underlying mechanism for rhythmic protein abundance in the SCN. Sol *et al.* [156] discovered a rhythmic stability for the *Per* transcript in *Drosophila*. Cheng *et al.* [157] found that the 3'UTR of this gene is alternatively spliced and results in two isoforms. Later, it was shown that the mode of splicing depended on temperature and photoperiod and contributed to locomotor-activity in long and hot days[158], which may help to avoid desiccation of the fly.

In subsequent studies, single core clock transcripts were the focus, with the aim to find the trans-acting factors. For the mouse transcripts of *per3* [159], *per2* [160] and *cry1* [161] rhythmic half-lives have been noted. In each of these cases the trans-factor, a known RBP, binds to the 3'UTR of the transcript in a rhythmic manner which leads to a time-dependent mRNA destabilization.

Another RBP known to affect the circadian clock is RBM4. RBM4, also known as LARK, has a circadian abundance in mouse SCN and *Drosophila* [162]. LARK is involved in diverse cellular processes that include alternative splicing of pre-mRNA, translation, and RNA silencing [163]. Altering the endogenous abundance of LARK in mice leads to an altered core clock period most likely by changing translational efficiency of the *per1* transcript [162].

Different deadenylases attack the poly(A) tail, a stabilizing element at 3'end of each mRNA, and thereby affect transcript stability and translational efficiency. Nocturnin, one of the mammalian deadenlysases, exhibits circadian rhythms at the transcript level in the retina of *Xenopus* [164] and in several mouse tissues with an especially high amplitude in mouse liver [165]. The targets of Nocturnin are not yet known [166]. However, Nocturnin knocked-out mice stay lean under a high-fat-diet indicating an altered lipid metabolism, accompanied by changes in glucose and insulin sensitivity [167].

With the advances of sequencing techniques it was possible to characterize the complement of RNAs bound by specific RBPs. To date this has been done for two RBPs with a circadian abundance, the cold-inducible RNA binding protein (CIRBP) and RNA binding protein 3 (RBM3). Both proteins are rhythmically expressed upon temperature cycles [168, 169] due to temperature-dependent splicing efficiency [170]. This expression behavior possibly enables

a synchronization of peripheral tissues via the body temperature. Identification of the binding sites of both proteins resulted in genome-wide listings of possible target mRNAs. More than 8500 binding sites for CIRBP and over 9500 for RBM3 were assigned to annotated transcript regions [169].

The importance of miRNAs for the circadian output is not clear. Gatfield *et al.* [171] investigated miR-122, a highly abundant, hepatocyte-specific microRNA [172]. Transcription of miR-122 is under circadian control, however due to a long half-live (see also Section 2.1) the miR-122 level is constant throughout the day [171]. Downregulation of miR-122 revealed that this miRNA affects a disproportionately large range of RNA with a circadian abundance and consequently, a knock-down results in reduction of lipid and cholesterol metabolism [173, 174], processes known to have diurnal rhythms. In line with this, a recent genome-wide study found 57 miRNA, all originating from primary transcripts of clock-controlled genes [175]. Consequently, the pre-miRNA levels oscillated, but the mature miRNA did not. An overexpression of one of these miRNA, miR-378, again revealed that this miRNA regulates many circadian genes.

Genome-wide screenings identified only a very small number of miRNA with oscillating abundances, six in *Drosophila* head [176] and 54 in mouse liver [129]. Du *et al.* [177] investigated the circadian transcriptome within a global disruption of miRNA biosynthesis. Here, 30% of circadian 1630 transcripts changed their mean abundance, however only 20 transcripts showed alterations in rhythmicity. Very few cycling miRNA and almost no alteration in mRNA abundance rhythms may indicate that miRNA generally do not induce circadian rhythms. This is supported by results showing that the kinetics of miRNA interactions are too slow to affect circadian rhythms in mRNA levels [178]. This is further supported by our study [179], where mRNA predicted to be under rhythmic post-transcriptional regulation show no over-representation of target sites for the rhythmic miRNAs suggested in above studies [129].

A very recent study suggests that long non-coding RNA may be involved in rhythmic post-transcriptional regulation. Torres *et al.* [180] investigated the content of paraspeckles, structures that are formed by a long non-coding RNA molecule and a number of RNA binding proteins [181, 182]. Paraspeckles are thought to prevent certain mRNAs from leaving the nucleus and, therefore, stop them from being decoded to proteins [183]. Both, the composition and the abundance of paraspeckles changed with a circadian rhythm in cells of the rat pituitary glands [180]. Disruption of the paraspeckles resulted in rhythm-loss of cytosolic mRNA levels known to be recruited by paraspeckles.

1.2.2 Investigation of Cis-Acting Elements

Charting of trans-factors is one way of investigation post-transcriptional regulation. However, insight can also be gained through methods: by investigating the modifications of cis-acting elements on the (pre)mRNA regardless of the mediating molecule. Such cis-acting motifs can be mediated by alternative splicing, poly(A) deadenylation and RNA methylation.

Alternative splicing in mammals is widespread, affecting 94% of genes in humans [184]. It does not only result in different proteins by, *e.g.*, exon-skipping, but can also affects the regulatory regions of mRNA, the 5' and 3'UTRs [185]. It is estimated that 20% of expressed genes in mouse liver contain a circadian exon [186]. McGlincy *et al.* [186] also demonstrated that certain splicing factors have clock-controlled transcript expression. Gotic *et al.* [170] observed that

temperature-dependent splicing efficiency of cold- and heat-inducible proteins lead to a circadian accumulation of mature mRNA. Only properly spliced mRNA can serve as a protein template, unspliced mRNA is degraded in the nucleus. Gotic *et al.* [170] found that a regulatory element in the first intron of Cirbp mRNA is responsible for changes in splicing efficiency.

The poly(A) tail is a stabilizing element of about 250 adenosides at the 3'end of mRNAs. Its length is connected to mRNA stability and translational efficiency. Once the mRNA is in the cytosol, the poly(A) tail length is attacked by different deadenylases [187]. But this shortening is not constant: It can slow down, stop, and even re-lengthening is possible to accommodate different mRNA fates [188]. Kojima *et al.* [189] investigated the poly(A) tail length of transcripts over the course of a day and find 230 transcripts with oscillating poly(A) tails. Surprisingly, most of them are not affected by a knock-out of Nocturnin, the only known rhythmic deadenylase [190].

m^6A RNA methylation is the post-transcriptional addition of a methyl group to an adenosine. Dominissini *et al.* [191] identified RNA methylation in more than 7000 human genes, typically around stop codons and within long internal exons. Fustin *et al.* [192] identified RNA methylation on many clock gene transcripts and demonstrated that upon inhibition of m^6A methylation the core clock period was longer.

1.2.3 Quantifying the Extent of Post-Transcriptional Regulation

If interested in rhythmic post-transcriptional regulation one may not simply ask and investigate the specific situations and players, but also how many of all expressed genes are affected by post-transcriptional regulation.

Since post-transcriptional regulation comes in many different flavors, as the previous section demonstrated, no single experiment can capture every possible post-transcriptional regulation at once. Hence, the problem is tackled the other way around by charting production and abundance of mRNA. If production and abundance "match" each other, there is no reason to assume rhythmic post-transcriptional regulation. For this approach it is crucial to find a clear and statistically sound definition of such a "match".

Charting of the mRNA abundance is achieved by common RNA-sequencing or hybridization-based ("micro-arrays" or "tiling arrays") studies. However, monitoring of production in a genome-wide manner is not as straight-forward. Different approaches have been developed, both experimental [193] and computational [70, 194], also employed to characterize the extent of rhythmic post-transcriptional regulation.

Experimental approaches to characterize transcriptional activity based on sequencing include ChIP-seq of polymerase II, GRO-seq and Nascent-seq. ChIP-seq (Chromatin Immuno-Precipitation & DNA-Sequencing) is used to examine the DNA binding sites of specific proteins. Proteins are cross-linked to DNA, followed by DNA fragmentation and immunoprecipitation by antibodies specific to polymerase II (Pol II). The sequencing of DNA fragments bound by Pol II allow the identification of Pol II enriched regions [195], however this covers all stages of transcription, including sites of Pol II pausing, where there is no active transcription. Global Run-On Sequencing (GRO-seq) [196] attempts to chart engaged polymerases. Here, nuclei with artificially halted transcription are isolated, the transcription is restarted *in vitro* and newly synthesized RNA fragments are isolated and sequenced. NascentSeq [197] is based on isolation of chromatin together with bound Pol II and transcribed RNA followed by RNA extraction.

This RNA is then depleted of poly(A) tails and rRNA and subsequently sequenced. Hence, all these three experimental approaches attempt to capture current transcriptional activity.

An established computational approach estimates RNA transcription from RNA sequencing data. Here, exons and introns are annotated separately accounting for mature RNA and pre-RNA, respectively. Exons reads can originate from both mRNA and pre-RNA, while intron reads can only stem from pre-RNA. However, most transcripts are spliced co-transcriptionally [198]. Hence, the complete set of introns is not part of the sequencing library. This leads to reduced mean level of pre-RNA quantification. A reduced mean level does not necessarily effect rhythmicity, and the data can still be used for qualitative interpretation (phase) of oscillations in transcriptional activity. However, to enrich for mRNA, the sequencing libraries are commonly generated via Poly(A) tails. These are added after completing the transcript. This means, firstly, that the moment we capture with this computational approach is after the completed transcription in contrast to the experimental approaches where transcription is "caught in the act". Secondly, due to co-transcriptional splicing transcripts may already be properly spliced and we cannot detect any pre-RNA levels for some transcripts. Despite these limitations, the quantification of pre-RNA and mRNA from one RNA sequencing run is an established method and seems to give useful insight also in data concerning the circadian transcriptome [70, 194].

To estimate the extent of rhythmic post-transcriptional regulation Koike *et al.* [70] and Menet *et al.* [5] divide their findings in 3 classes of genes: first, genes with rhythmic transcriptional activity and rhythmic mRNA abundance; second, genes with rhythmic transcriptional activity but an arrhythmic or flat mRNA abundance; and third, genes with arrhythmic transcriptional activity, but rhythmic mRNA abundance. They conclude, that the first class (rhythmic-rhythmic) is not under control of rhythmic postranscriptional regulation; in the second class (rhythmic-arrhythmic) rhythms are lost due to long mRNA half-lives; and only the third class (arrhythmic-rhythmic) contains genes under rhythmic post-transcriptional control. Although not completely wrong, these conclusions do not hold for every examined gene. For example, even if both, transcriptional activity and mRNA abundance are rhythmic, there are cases where transcriptional activity cannot explain mRNA abundance, for example if the phase difference between both is large or mRNA abundance oscillates with a larger relative amplitude than the corresponding transcriptional activity. This can be comprehensively understood using a simple ordinary differential equation model, see also Section 2.1. The detailed analysis of Le Martelot *et al.* [6] is based on this model.

In our study [179] we reexamined the data of Le Martelot *et al.* [6]. We developed a statistical test to precisely estimate how likely it is for a given mRNA to be under rhythmic post-transcriptional control. The main advantages of this test is, firstly, to overcome potentially misleading assignments of binary categories, either rhythmic or arrhythmic, and secondly, to handle every gene separately whilst dealing with its specific measurement noise. Using this test we were able to estimate that about one third of the circadian transcriptome in mouse liver (data from [6]) is under rhythmic post-transcriptional control. A very similar result was found for *Drosophila* head (data from Rodriguez *et al.* [199]) where 34% of circadian genes showed evidence of rhythmic post-transcriptional control.

In the last section of my introduction, I will briefly outline the ideas behind this modeling approach. This thesis is based mostly on these ideas, as published in [179], but I will present

significant extensions of the model and show its versatility by applying it to a series of hypotheses on rhythmic regulation on various of above outlined levels in the complex life of genetically encoded macromolecules.

1.3 Modeling Idea and Detection of Rhythms

In this last section I will briefly comment on the modeling technique I have implemented. First of all, I concentrate on describing deterministic behavior and, with that, exclude any stochastic variation. Almost all models described in this thesis, except for the partial differential equation from section 3.3, are compartment models with time-dependent rates.

The compartments emplyed are not physical spaces but pools of molecules. With applying compartment models I assume that

- There is an instant homogeneous distribution of molecules within a "compartment."
- The cell volume is constant, that is, effects of cell growth are not accounted for.

In many cases I describe only a single compartment with a time dependent input, molecule production, and time dependent output, generally molecule degradation. This rather simple description does not account for the underlying biochemistry, for example molecular interaction, inhibition *etc.*, in any detail. This makes the models universally applicable. Although I focus on RNA production and degradation, many conclusions hold true for other molecules, *e.g.* proteins, orthophosphate proteins, nuclear import/export, etc., see Section 4 for more examples.

1.3.1 Rhythms in Time Series and Rates

Any sustained oscillation can be characterized by four features, see also Figure 1.3A:

- magnitude: the mean value
- period: the time it needs to complete one oscillation
- phase: the peak (or trough) of an oscillation with reference to an external point such as time
- amplitude: the difference between peak (or trough) and magnitude

For circadian behavior the period is fixed to about 24 hours and will not be of great interest. In this thesis I also will rarely mention the mean value since I will not concentrate on comparison of absolute expression levels but on comparison of their dynamics, their rhythmicity. Of specific interest are the phase of the signal (when does molecular abundance or activity peak) and amplitude (how strong does it vary over one cycle).

Here, it is convenient to introduce a normalized amplitude, the so-called relative amplitude. The dimensionless, relative amplitude is defined by the absolute amplitude normalized by the mean and is a measure for rhythmicity. It can take values between 0, no rhythmicity, and 1, maximally possible oscillation.

The rhythms in time series and rates are then characterized by a phase and a relative amplitude.

I will describe most time-dependent rhythmic rates $r(t)$ as cosine functions, if not stated otherwise, with magnitude M, phase ϕ and relative amplitude A_{rel}, see Figure 1.3A.

$$R(t) = M\,(1 + A_{\text{rel}}\cos(\omega t - \phi)) \tag{1.1}$$

The angular frequency is set to $\omega = 2\pi/(24\text{h}) \approx 0.26\ \text{h}^{-1}$ throughout this thesis to reflect circadian biology. The extreme values of the relative amplitudes, 0 and 1, indicate for 0, rate is constant, and 1, rate oscillates between 0 and twice its mean.

1.3.2 Vector Model to Describe Rhythms

To visualize rhythmic properties of cosine functions it is convenient to use a vector description. Here, one vector represents one cosine function. Its length represents the oscillation strength or relative amplitude and the vector's direction the oscillation phase, see Figure 1.3B. If we want to describe processes with a circadian rhythm the phase can take values between 0 and 2π, which represent values between 0 and 24 hours in the circadian time frame. Although, all calculations are done in angular frequency I will display and write about the corresponding values in the circadian time frame. As with the rates also rhythmic time series can also be approximated with cosine functions. Here, a linear model, *e.g.*

$$x(t) = a\cos(\omega t) + b\sin(\omega t) \tag{1.2}$$

is fitted to a time series, estimating the parameters a and b. Since both models, Equation 1.1 and 1.2 are equivalent, we can calculate from the parameters a and b phase ϕ and relative amplitude A_{rel} with

$$A_{\text{rel}} = \sqrt{a^2 + b^2} \tag{1.3}$$

$$\phi = \arctan 2\left(\frac{a}{b}\right), \tag{1.4}$$

where the function arctan2 maps the two arguments a and b onto the whole phase plane, see Appendix A.1.2. With this the time series is described as a cosine function and can be displayed as a vector as previously described. The derivation between fit and time series, the residuals of the fit, are displayed as a so-called confidence ellipses in this graph, see Figure 1.3C. For most experimental designs, namely fully measured periods, the ellipse is a circle, see Appendix C.1.3. The circle covers a very small proportion of possible phase values compared to possible amplitude values. This has the consequence that phase can be estimated by much greater accuracy than the relative amplitude.

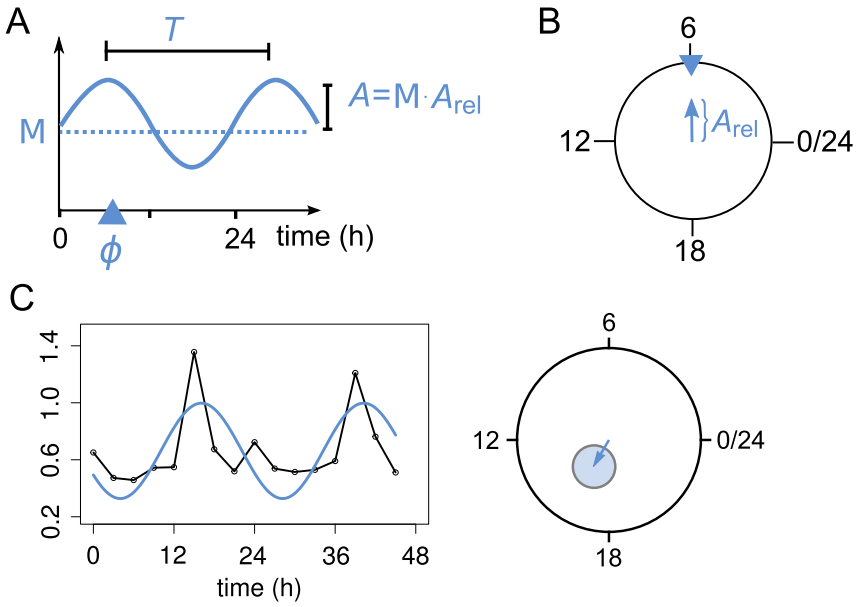


Figure 1.3: **Definition and depiction of rhythmic properties.** **A** A cosine oscillation is characterized by its period $T = 2\pi/\omega$, magnitude M, phase ϕ and amplitude A. The relative amplitude A_{rel} is defined as the amplitude normalized by the magnitude. **B** Cosine oscillation of panel A is represented as vector where the length of the vector accounts for the relative amplitude, the direction for the phase. **C** A time series is fitted with a cosine-function. This cosine is again represented as vector, residuals from the fit are represented as 95% confidence ellispe, see Appendix A.1.2.

1.3.3 Statistics of Rhythmicity Detection

Numerous high-throughput studies on the circadian transcriptome obviously ask for the number of cycling genes compared to the overall expression. Often it is implicitly assumed that rhythmicity in transcript abundance implies rhythmicity in biological function. To be biologically functional rhythms should exhibit a sufficient amplitude, oscillations must be strong. In contrast, when, for example, asking for principal regulatory mechanisms of the circadian clock one might also be interested in rhythmic gene expression with low amplitude regardless if the rhythm has a function or not. Hence, different scientific questions pose different demands on rhythmicity detection. I argue it is useful to test for rhythmicity and amplitude independently, although both are related, as I will outline in the following. With this I follow the suggestion by Thaben *et al.* [200].

To find circadian rhythms in time series often a statistical hypothesis testing is employed which tests for rhythmicity against a null hypothesis as some sort of random noise without any rhythmicity. There are statistical tests based on both parametric [201, 202] and non-parametric methods [203, 204, 200] available, reviewed in [205], with non-parametric often being more robust against biological outliers.

In the literature, fixed percentages or numbers of genes are often given, as I also reported in Section 1.1.5. These are almost always based on applying a p-value cut-off of *e.g.* 0.05, chosen arbitrarily. However, testing thousands of genes for rhythmicity is a multiple testing problem and hence asks for correcting the p-values, for example by using the Benjamini-Hochberg procedure [206].

In this thesis, in order to detect rhythms in time series I apply RAIN [200] together with a false-discovery rate of 25%. RAIN is a further development of a non-parametric method to detect not only symmetric, but also asymmetric wave-form rhythms. This yields very high numbers of rhythmic transcripts, *e.g.* applying this method to the liver data set by Zhang *et al.*[94], we find that at least 49% of transcripts are rhythmic, much more than the published number of 16%.

However, most of these detected rhythms have a very low amplitude. Hence, this detection method is accompanied by an amplitude cut-off. This amplitude cut-off can be motivated biologically. For example, as it is well established and as I also outline in Section 2.1, rhythms are lost easily due different processing steps. Accordingly, if I want to deduce from rhythmicity in transcript abundance to rhythmicity in protein I should choose a high amplitude cut-off. On the other hand, if I am rather interested in basic regulatory principles in RNA processing, a low amplitude cut-off is sufficient.

Using this approach to search for rhythmic transcripts I first find a reliable set of cycling time series and then chose a motivated, in contrast to arbitrary, amplitude cut-off which serves best for my scientific question.

1.4 Outline of the Thesis

This thesis aims to elucidate the effect of post-transcriptional, especially rhythmic post-transcriptional regulation on the circadian transcriptome. How can the circadian clock tune its output to specific times and keep a desired oscillation strength? I investigate the main properties of

rhythmicity, phase, relative amplitude and period.

First, in Section 2.1 I look at constant, *i.e.* not time-dependent, post-transcriptional regulation, and how it affects amplitude and phase of a transcript. This is the starting point to developing different tests which estimate the extent of rhythmic post-transcriptional regulation in Section 2.3 and 2.4. I apply these tests to two data sets on mouse liver and kidney, revealing 34% of circadian genes in mouse liver and 18% in mouse kidney for which a rhythmic post-transcriptional regulation must be assumed. In Section 2.5 I introduce a model for rhythmic mRNA degradation and show that rhythmic degradation could account for the measurements. In Section 2.6 I ask if rhythmicity in other stages of mRNA processing could equally explain the data. In Section 2.7 I finally search for additional evidence of rhythmic post-transcriptional regulation in mouse liver and kidney.

The second part of this thesis complements this general investigation of the prevalence of multiple rhythmic processes with more specific questions. First, I ask if and when higher harmonics in time series could be induced by rhythmic post-transcriptional regulation (Section 3.1). I further inspect rhythmic post-transcriptional regulation of core clock genes in Section 3.2. Finally, in the last section I introduce a partial differential equation model, which describes an "aging" molecule. This model is applied to damaged proteins and shortening of poly(A) tails.

2 Post-Transcriptional Regulation of Clock Controlled Genes

The transcriptome of clock controlled genes in different mouse organs show a phase distribution throughout the day. It seems different genes require different phases in order to properly fulfill their function. This requires for a fine adjustment of phases and amplitudes of mRNA abundances. Transcript levels are tightly regulated via different steps, starting with the transcription initiation, followed by splicing and other post-processing and finally the degradation of the mature transcript. Genome-wide diurnal profiles of mRNA abundances are obtained by next generation sequencing technology. A genome-wide approach for transcriptional activity can be achieved by different techniques such as "nascent-seq" [197], "GROseq" [207], ChIPSeq of Polymerase II or identification of introns in RNA sequencing data. However, for the diverse range of most PTR steps a genome-wide approach has not yet been established.

In this work I aim to elucidate the importance of post-transcriptional regulation (PTR) for circadian gene expression with a focus on rhythmic PTR. With ordinary differential equation (ODE) models I will characterize the limits of constant PTR, how rhythmic PTR affects mRNA levels and how rhythmic PTR can be detected without direct experimental evidence. I will use 2 data sets on mouse liver and kidney to test this. In the second part I will extend these models to characterize other interesting aspects of rhythmic PTR. I will investigate if rhythmic PTR can produce higher harmonics in mRNA abundance, briefly characterize rhythmic PTR in the core clock and lastly describe the life cycle of a molecule as an "aging" process by a partial differential equation (PDE) model.

Part of this work has been published in "Rhythmic degradation explains and unifies circadian transcriptome and proteome data" [179], in particular the solution and stability discussion of the model in Section 2.5, and the test which uses half-lives presented in Section 2.4. This test I developed in collaboration with Paul Thaben and Pål Westermark. Apart form this, the work I present here is completely my own.

2.1 Time Matters

The kinetics of mRNA processing are crucial for generating rhythms in mRNA abundance. They dictate strict boundaries for maintaining oscillations. In the following I will use an ODE model to show what determines the oscillation strength (amplitude) and how oscillations are passed during mRNA processing steps.

Rhythmicity or oscillation refers to relative amplitude. The relative amplitude defines how strong the rhythm is. A loss in amplitude means a loss of rhythmicity.

How is a rhythm in transcription initiation passed on to the mature transcript? Since we are

interested in oscillation only, we can describe several life stages of an mRNA by one model equation. This is then a global description without taking into consideration underlying molecular mechanisms. In this equation a species x is rhythmically produced and constantly degraded:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \mathrm{prod}(t) - \gamma x \\ \mathrm{prod}(t) &= k\left(1 + A_{\mathrm{prod}} \cos\left(\omega t - \phi_{\mathrm{prod}}\right)\right). \end{aligned} \tag{2.1}$$

The time dependent production rate $\mathrm{prod}(t)$ is a cosine function with period $\omega = \frac{2\pi}{24}\mathrm{h}^{-1}$, mean k, relative amplitude A_{prod} and phase ϕ_{prod}. γ is a constant degradation rate characterizing the removal of the molecule out of the described system. This rate is related to the species' half-life or processing time $\tau_{1/2}$ by the relationship

$$\gamma = \frac{\ln(2)}{\tau_{1/2}}. \tag{2.2}$$

The different interpretations for the different life stages are achieved by interpreting the variable and rates differently. For example, let x stand for *nascent, unspliced RNA*. Then the rhythmic production rate describes a rhythmic transcription initiation. In this interpretation, splicing destroys nascent RNA, hence the splicing rate constitutes the degradation rate. The degradation rate is defined by the processing time, here, how long it takes for splicing. The subsequent species *spliced RNA* is rhythmically produced due to rhythmic pre-RNA abundance, the pre-RNA passes its rhythm down to the spliced RNA. In the same manner every life stage of an mRNA is described by the same model, where the previous species is approximated by a rhythmic production rate.

Solving the model Equation 2.1 gives a cosine function describing the abundance of x with magnitude, relative amplitude and phase:

$$x(t) = M_x\left(1 + A_x \cos\left(\omega t - \phi_x\right)\right), \tag{2.3}$$

$$M_x = \frac{k}{\gamma}, \tag{2.4}$$

$$A_x = \frac{\gamma A_{\mathrm{prod}}}{\sqrt{\gamma^2 + \omega^2}}, \tag{2.5}$$

$$\phi_x = \phi_{\mathrm{prod}} + \arctan\left(\frac{\omega}{\gamma}\right). \tag{2.6}$$

According to this solution, the relative amplitude of x is determined by the relative amplitude of the production, reduced by the factor $\frac{\gamma}{\sqrt{\gamma^2+\omega^2}}$. Consequently, the processing time or species' half-life in the described system dictates the strength of amplitude reduction. Accordingly, every processing step of the mRNA will result in an amplitude reduction in its abundance or diminishes oscillation strength. Figure 2.1 shows the amplitude reduction in relation to the rhythm's period. For example, the species' oscillation is reduced by one half if the processing time takes about 20% of the period.

Which steps in mRNA processing will therefore contribute most to an amplitude reduction in a circadian time frame where the period equals 24 hours? In Figure 2.1 the time scales of some mRNA processing steps are plotted against the amplitude reduction. The life of an mRNA begins with its transcription initiation. Assuming transcription speed is ∼35 base per second as measured by Bahar *et al.* [208], the completion of a transcript with a length of 100kb takes 45 min. Further processing of the freshly transcribed mRNA includes splicing, adding its stabilizing elements, poly(A) tail and the 5'cap, and finally transport out of the nucleus. Splicing of one intron takes about 10 minutes [209]. However, since most splicing occurs co-transcriptionally [198] this step will not prolong the processing time of mRNA during transcription. The same is true for the co-transcriptional mRNA capping. Adding the poly(A) tail, in mammals consisting of 250 adenosides, takes no longer than 5 minutes [210]. Nuclear export can take up to 30 minutes depending on where in the nucleus the transcript is produced [211]. In total, the time from transcription start to the nuclear export can take up to 1.5 hours. Hence, an oscillation in transcription initiation, *e.g.* by an oscillating transcription factor, is already reduced by 40% due only to the long mRNA processing time.

However, 1.5 hours is quite an extreme example for mRNA processing. In most cases the processing, from the start of transcription until nuclear export, will probably not exceed 0.5 hours due to shorter transcripts and faster nuclear export and hence this step will not play the dominant role in rhythm loss. By far the most amplitude reduction can be expected due to the mRNA's half-life. Friedel et al [212] and Schwanhäusser *et al.* [213] measured half-lives of mRNAs in mammalian cell lines. Both studies used labeled nucleotides to distinguish newly transcribed from pre-existing RNA. Half-lives were then estimated by the ratio of both fractions. According to these studies, mRNA half-lives span a range from a few minutes to up to 25 hours, with a median of 5 hours. Hence, for long-lived mRNA the oscillation of its production will be lost even if production strongly oscillates.

Long mRNA half-lives weaken rhythms. Furthermore, the half-life also influences the phase of the mRNA. It is intuitively clear, that a product abundance peaks after its production, the abundance of a mature mRNA will peak after its rhythmic transcription initiation. The solution of the ODE Equation 2.6 gives a more precise time frame. The product abundance peaks after its production at most at 1/4 of the period, 6 hours for the circadian period. The longer the half-life the longer is the phase shift.

In summary, it depends mainly on the mRNA half-life if an mRNA has a rhythmic abundance due to a rhythmic transcription initiation. Long half-lives result in a loss of rhythms accompanied with a phase shift of up to 6 hours between transcription initiation and mRNA abundance. This result also gives a good starting point to characterize 2 RNA sequencing data sets with additional information about transcriptional activity. In the next section, I will introduce both data sets in more detail.

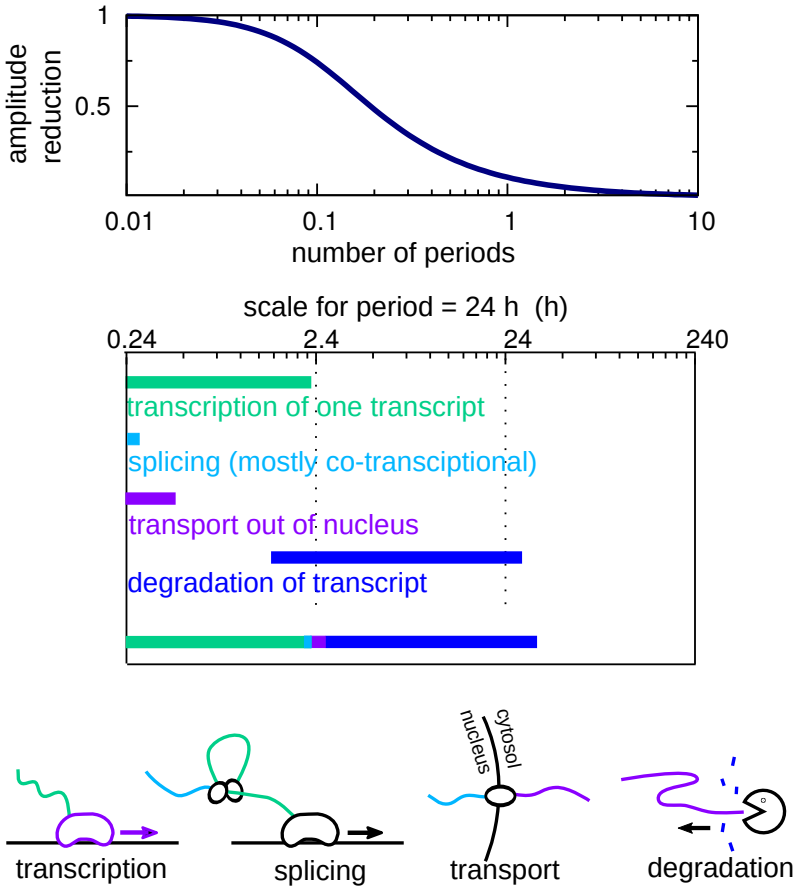


Figure 2.1: **Time matters for oscillations.** Above: Amplitude reduction in percentage plotted against the processing time in number of periods. Below: Range of different mRNA processing steps.

2.2 RNA Abundance and Transcriptional Activity: 2 Data Sets on Mouse Liver and Kidney

All further analyses in the following sections will examine two data sets. In this section I will first characterize the properties regarding circadian gene expression. The first data set is derived from mouse liver and is published by Menet *et al.* [5], the second data set was generated from mouse kidney by our group in collaboration with the group of Achim Kramer (Charité) and Roman-Ulrich Müller from University of Cologne. Both data sets are based on RNA sequencing. To access transcriptional activity, the second feature of the data sets, different methods are utilized. Menet *et al.* [5] make use of the so-termed "nascent-seq" method, explained in Section 1.2.3. Here, they separate experimentally pre-mature mRNA from mature mRNA followed by sequencing. In contrast, transcriptional activity in kidney is obtained from the same RNA sequencing data set which provides mRNA abundance. Separation between transcriptional activity and mRNA abundance is performed computationally by individually quantifying transcripts with (pre-RNA) or without (mature mRNA) introns, a method which possibly provides not as "clean" results as the method applied by Menet *et al.* [5] as described earlier. Furthermore, the sampling frequency is different between the two data sets. Mouse liver was sampled every 4th hour, 6 time points per period and 12 time points in total, mouse kidney was sampled with a slightly higher frequency, every 3rd hour, 8 time points per period and 16 in total.

In the following 3 sections, there will be a lot of numbers arising regarding these two sets. For the convenience of the reader these numbers are summarized in a table in Appendix F.2.

In mouse liver 13698 genes are expressed with information on both, mRNA and transcriptional activity, in kidney 14324 genes are expressed. I consider circadian genes as genes with mRNA or transcriptional activity with a 24 hour-periodic pattern (detected with RAIN [200] and a false discovery rate ≤ 0.25) and a relative amplitude larger than 0.1, see also Section 1.3.3. With these cut-offs I find 3813 (30%) of expressed mRNA to be rhythmic in liver and 4137 (29%) in kidney. If I include transcriptional activity into this analysis I find that 5581 (43%) of expressed genes have a circadian rhythm in either their RNA abundance or transcriptional activity, for kidney this is true for 6489 (47%) of expressed genes. Compared to other studies these percentages of circadianly expressed genes are more than twice as high [94, 5]. This can be in part explained by a higher sampling frequency and the different experimental methods. RNA sequencing employed in the present studies produce data less prone to technical noise compared to microarrays [214]. Both, sampling with higher frequency and less noise, increases the detection of rhythms. Furthermore, RAIN, not employed by the other studies, is able to detect rhythms which other known detection algorithms miss [200]. However, this is not enough to explain the large discrepancy, also the parameter thresholds which separate circadian from non-circadian gene expression play an important role. Here, rather the choice of relative amplitude than a different false discovery rate affects the proportion, see Appendix C.1.2. Hence, many genes I classify as circadian have a low relative amplitude, see also Figure 2.2A. One may argue, that a large proportion of these genes are not able to fulfill a circadian function in the cell. However, the main purpose of this study is to investigate rhythmic PTR. Not asking for specific gene functions but rather for fundamental regulation principles justifies rather loose parameters for gene classification in order to keep the test set as large as possible.

A different phase distributions of mRNA abundance and transcriptional activity for liver and kidney suggest that circadian gene expression is organ specific, see Figure 2.2B. A clear difference between both organs is also found for the phase difference of transcriptional activity and mRNA abundance, see Figure 2.2C. In kidney most of the mRNA abundance peaks within 6 hours after its transcriptional activity, while in liver a broader distribution of phase differences is observed. The theoretical boundary for a phase difference is 6 hours when rhythms in mRNA abundance stem only from rhythmic transcriptional activity, see previous section. Hence, one would expect the proportion of genes where rhythms in mRNA are only generated by a transcriptional activity to be much higher in liver than in kidney. However, this needs further validation.

Consistent with the organ specific circadian transcriptome, I find that most of the genes (12148) are expressed in both organs, but only a small proportion (1361 genes, 11% of genes expressed in both organs) is found to be circadian in both organs.

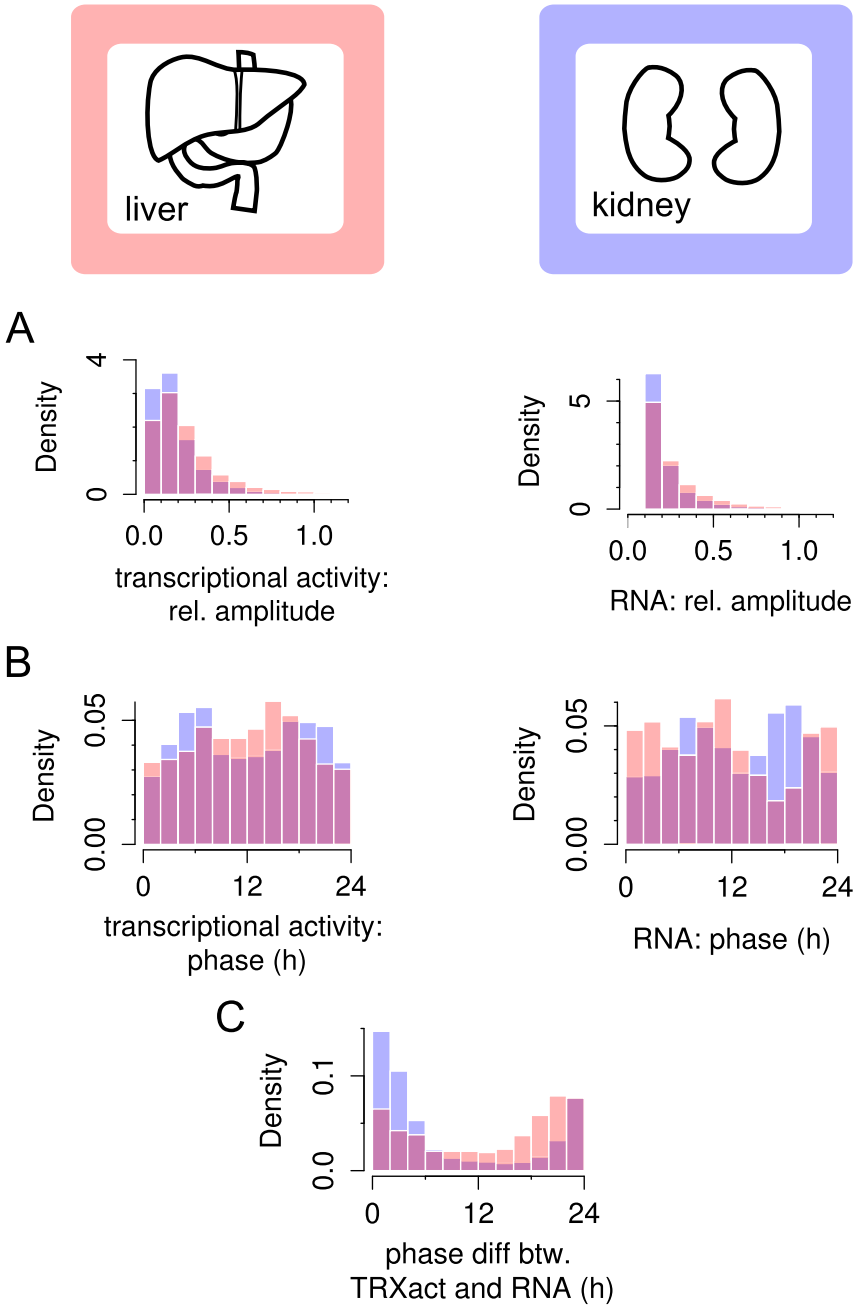


Figure 2.2: **Characteristics of circadian gene expression in liver and kidney.** Shown are different distributions. In order to compare both organs, not absolute numbers but the densities are shown. Transparent red: liver, transparent blue: kidney. The overlay of both appears violet. **A** Relative amplitude of transcriptional activity and mRNA abundance. **B** Phase of transcriptional activity and mRNA abundance. **C** Phase difference between transcriptional activity (TRXact) and mRNA abundance ($\phi_{\text{TRXact}} - \phi_{\text{mRNA}}$). 0 refers to the phase of transcriptional activity.

High relative amplitudes suggest that rhythmicity is important for gene function. With that in mind it is not surprising that genes with a high relative amplitudes in their mRNA abdunance in both kidney and liver are almost exclusively core clock genes, see Figure 2.3A. For some other genes no connection to the circadian clock has yet been investigated, despite their strong rhythmicity, see Figure 2.3A. It seems that, to date, we have only scratched the tip of the iceberg when it comes to knowledge of output and consequences of the circadian clock. Furthermore, there are mRNA which show a relative amplitude larger than 1, which would include theoretically (see Section 1.3.1) negative RNA abundances. However, the time series of these mRNAs have a distinct, rather pointy shape and a sine fit results in larger relative amplitudes, see Figure 2.3C. It might be desirable to introduce a different fit to these time series. This is beyond the scope of this thesis.

The phases of genes expressed in both organs often differ among organs, see Figure 2.3B. In contrast, phases of genes with a high relative amplitude correlate quite well, see Figure 2.3B. Interestingly however, we observe a systematic phase shift between those genes whose phases correlate. Transcript abundances peak later, while transcriptional activities peak earlier in kidney than in liver. For the latter, the experimental setup may be blamed. Nascent-seq applied for transcriptional activity in liver captures a broad range of newly synthesized RNA including already spliced RNA. The computational separation of exons, proxy for the mature mRNA, from introns accounting for transcriptional activity captures only unspliced RNA. Since RNA is often spliced very early in its life time, the phase of transcriptional activity in kidney appears earlier than in liver.

However, the systematic phase difference in mRNA abundance - especially in core clock genes - is rather odd. It could mean that the core clock oscillates with a phase difference of about 5 hours in both organs. But we must take into consideration that the two data sets were generated in different laboratories, each with its own routines, mouse strains etc.. Although both experiments used mice housed in LD 12:12, probably the most significant influence on possible phase differences, it would still require an investigation of both organs in one laboratory in order to exclude any other experimental influences.

Two organs, two distinct circadian transcriptomes with some similarities, especially in core clock gene expression, leads to the question: How much of the circadian transcriptome can be explained by the transcriptional activity? There seems to be a difference between the two organs because many more mRNA in liver in comparison to kidney have phase differences between transcriptional activity and mRNA abundance larger than 6 hours, the theoretical boundary for constant PTR, see also Figure 2.2C. In the following section I examine both data sets by comparing transcriptional activity and transcript abundance with respect to the model from Section 2.1.

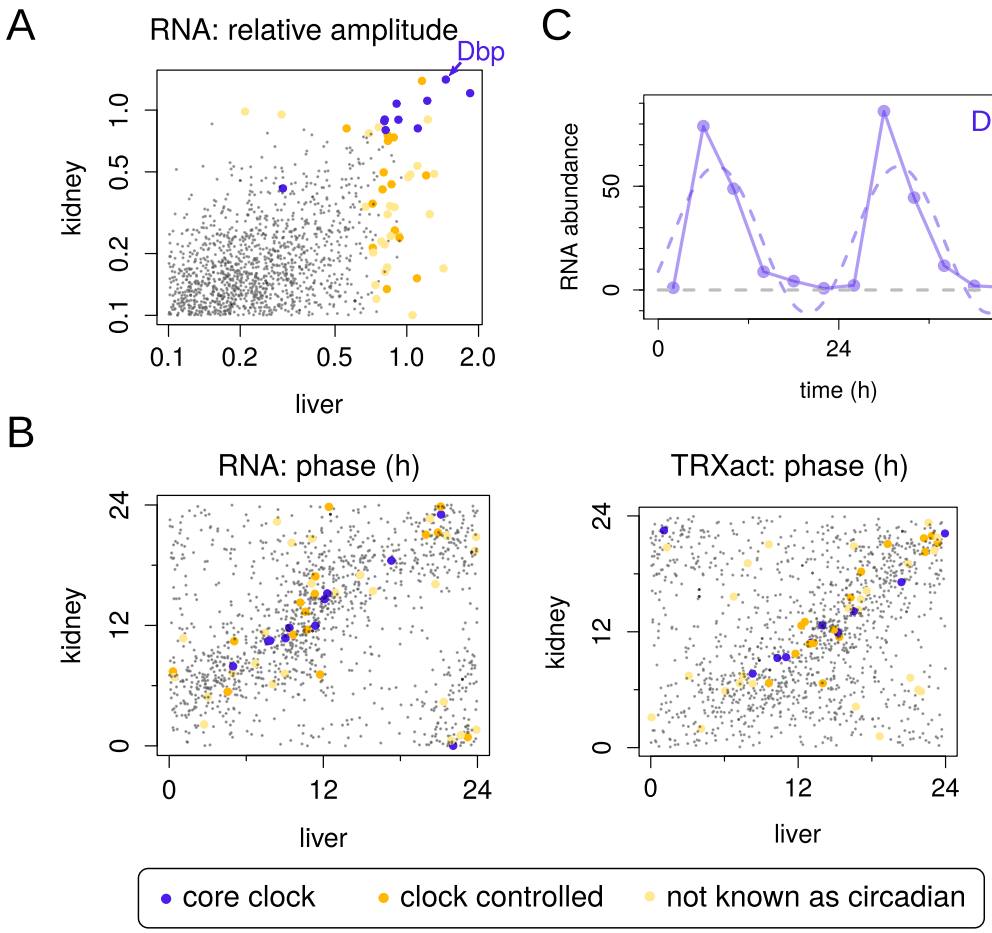


Figure 2.3: **Genes expressed both in kidney and liver. A** Relative amplitude of expressed genes in liver and kidney. Color coded are core clock genes (violet) (for list see Appendix F.1) or genes with high (>0.8) relative amplitude. Orange: pubmed-search "circadian + [gene name]" gave result, yellow: the same pubmed-search gave no result. **B** Data and fitted sine curve for Dbp demonstrating a relative amplitude larger than 1. **D** Scatter plot of phases for transcriptional activity and mRNA abundance in both organs. Color code same as in A.

2.3 Rhythmic Transcriptional Activity Cannot Fully Explain Rhythms in mRNA Abundances

How many of all rhythmically expressed genes can be explained by rhythmic transcriptional activity alone? I will develop in the following chapter a 2-stage-test which tests first mRNAs without half-lives. In the second, refined test stage half-lives are included. In this way, more genes can be tested.

The overarching strategy in both test-stages is to compare the model prediction from the previous Section 2.1, where only rhythmic transcriptional activity was considered, with the actual findings. If prediction and findings agree the mRNA abundance can be explained by only transcriptional activity.

2.3.1 A Test in Two Stages - Without and With Half-Life

To make full use of the model Equation 2.1 one needs three ingredients: time series of mRNA abundances, transcriptional activities as proxy for the production rate and the mRNA half-lives. However, information on only mRNA abundance and transcriptional activity allows a first estimate on the extent of rhythmic PTR as I will outline in the following. This estimate is not as precise as if mRNA half-lives were included, but it takes more transcripts into account since the majority of transcript half-lives were not measured and a more precise analysis is not possible for these transcripts.

The graphical representation of the model solution gives an intuitive understanding of the comparison between model prediction and measurement. As outlined in Section 1.3.1, every oscillation can be represented by a vector, where the vector direction indicates the oscillation phase and the vector length the oscillation strength or relative amplitude. The model solution predicts that the mRNA abundance peaks at most 6 hours after the transcriptional activity and is accompanied by an amplitude reduction. Exact phase shift and amplitude reduction depends on the transcript's half-life. This result translates into a semicircle in the vector-representation, see Figure 2.4A. The upper edge of the semicircle marks the phase shifts and amplitude reductions for a transcriptional activity with relative amplitude of 1, the highest possible relative amplitude. Consequently, this semicircle marks the maximal range for rhythmic transcripts which are only influenced by a rhythmic transcriptional activity. Transcripts influenced by transcriptional activity with a smaller amplitude lie within the semicircle. Transcript abundances outside of this semicircle cannot be explained by the model, indicating that there is an additional rhythmic process which influences the transcript abundance. These two fractions of mRNA rhythms, lying either inside and outside the semi-circle, allow for a first distinction between transcripts with or without additional rhythmic PTR regardless of their half-life.

Including the half-life allows for a further refinement of this distinction. For all transcripts within the semi-circle and for which the half-life is known, we can check if phase and amplitude match the model prediction.

This two-stage-test, first to test all transcripts regardless of their half-life and second further validation of phase and amplitude, will give an estimate for the extent of rhythmic PTR taking as much information as possible into account.

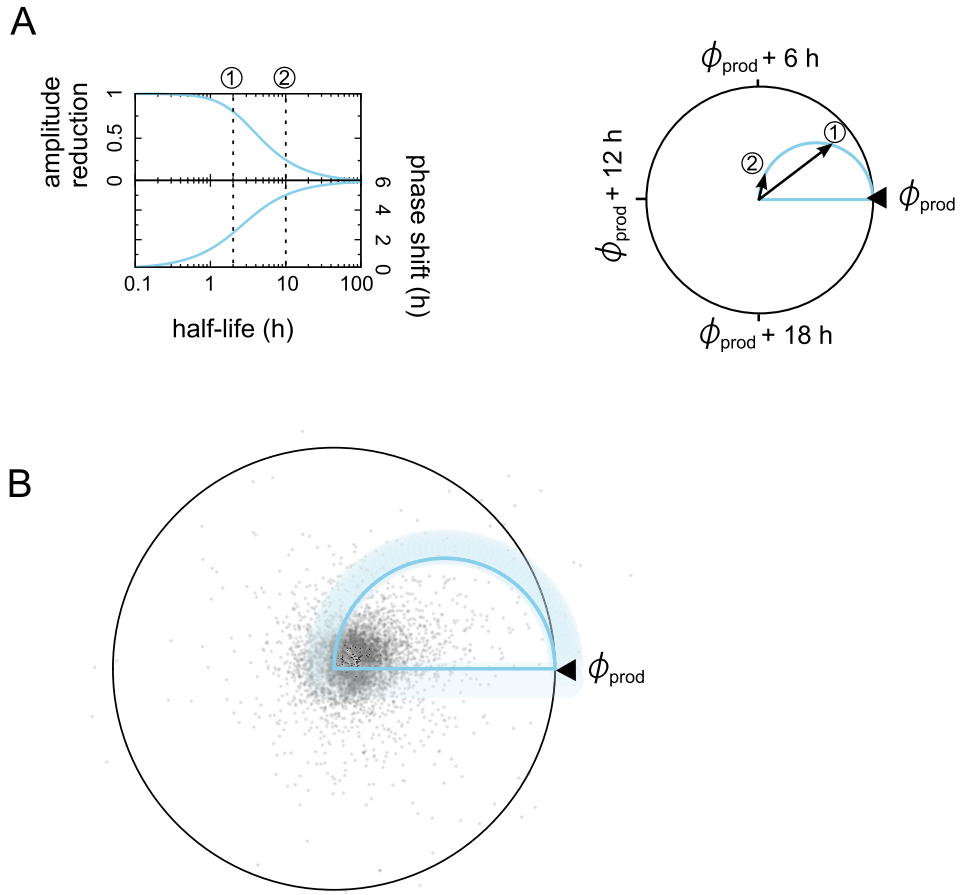


Figure 2.4: **Test for agreement of model prediction Equation 2.5, 2.6 and data in 2 stages: Principle and first test stage for kidney. A** Amplitude reduction and phase shift due to mRNA half-life as graph (right) and as vector representation (left). Numbers indicate two examples in both descriptions. **B** First test-stage for kidney data set. Shown are all mRNA abundances in the vector representation from A aligned to one transcriptional activity phase with $\phi_{\mathrm{RNA}} - \phi_{\mathrm{TRXact}}$, corresponding to $\phi_x - \phi_{\mathrm{prod}}$ in the model description. mRNA abundances outside the semi-circle (+ 0.1 size of semi-circle, indicated by shaded area) are considered as "cannot explained by their transcriptional activity".

2.3.2 Two-stage-test on Mouse Liver and Kidney

To apply the first stage of the test, namely to test if transcripts are inside or outside the semicircle, I align all mRNA phases to one production phase and allow the semicircle to be larger than dictated by the model, which takes uncertainties in amplitude and phase estimation into account, see Figure 2.4B for kidney.

With this set-up I find that of all rhythmic transcripts 939 (∼14%) are outside the semicircle in kidney and 2191 transcripts (∼36%) in liver, see Figure 2.4B.

The second stage of the test is applied to the remaining fraction, namely transcripts inside the semi-circle, for which a half-life was measured by Friedel *et al.* [212] and/or Schwanhäusser *et al.* [213]. That is the case for 1163 transcripts in kidney and 850 transcripts in liver. With the model solution Equation 2.6, 2.5 I calculate the expected phase and amplitude of transcript abundance based on the transcriptional activity and half-life. Next, these results are compared to the transcript phases and amplitudes that were actually measured and I allow again for a certain deviation to account for uncertainties in phase and amplitude estimation, see Figure 2.5 for kidney.

Here, a large part of the examined transcripts fail to pass the second test and hence their rhythmic abundance profiles cannot be explained by rhythmic transcriptional activity alone. In kidney this comprises of 483 transcripts (∼42%) and in liver 419 transcripts (∼49%). It seems that in many cases relative amplitudes are larger than the model prediction, see Figure 2.5. This may indicate that PTR generally leads to an increase in relative amplitude of mRNA abundance. Additionally, PTR might fine-tune phases, *i.e.* the timing of abundance peaks, as the differences between the phase distributions of model prediction and measurement suggest.

In summary, I could quantify in a two-stage-test that the majority of rhythmic transcripts cannot be explained by rhythmic transcriptional activity alone. The nature of the test, that is, to first test data on transcriptional activity and transcript abundance regardless of their half-life and second, to test the remaining fraction with known half-lives again, gives only a relative number of total transcripts under rhythmic post-transcriptional control. This is obtained by summing results from both stages. Accordingly, in the transcriptome of kidney more than half, 56% of all rhythmic transcripts, are under rhythmic-post-transcriptional control, in liver this fraction is with 85% extraordinary high.

These percentages are surprisingly high. Consequently, one may question the reliability of this result. Each test result depends on a binary decision, namely, does the oscillation belong to an allowed range or does it not. Uncertainties in phase and amplitude determination are generally handled as if uncertainties would be the same for all measured transcripts. Choosing a different threshold which allows a greater divergence from the model prediction would greatly influence the fraction of genes which cannot be explained by their transcriptional activity alone. For example, if I allow for a divergence from the model prediction by 30% instead of 20%, see Figure 2.5, the gene fraction which cannot be explained by their transcriptional activity decrease to 31% in kidney and 53% in liver. Such a strong dependence of the result on an arbitrary threshold is not desirable. Furthermore, there is transcript specific information on the reliability of phase and amplitude estimation for each transcript available, hence there is no need to choose one threshold valid for all genes.

However, no matter how trustworthy the fraction of transcripts with rhythmic PTR is, the

result demonstrates one fact: the observed rhythms in mRNA abundance cannot be generated by rhythmic production alone. To alter propterties of rhythm such as phase and amplitude it needs the influence of other rhythms. When a rhythmic transcriptional activity does not explain the rhythm in transcript abundance then there must be a rhythmic post-transcriptional process involved.

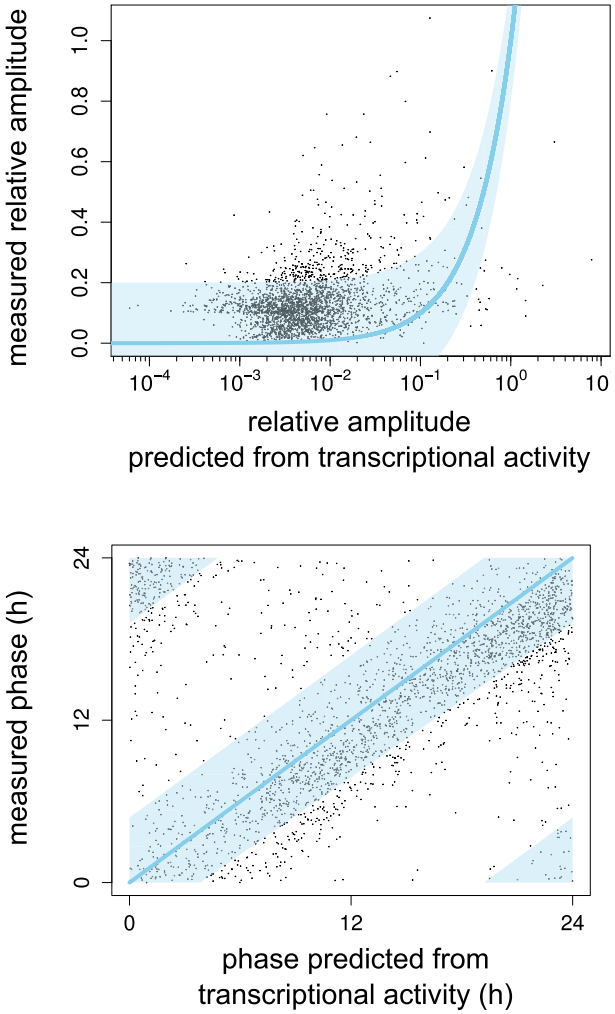


Figure 2.5: **Test for agreement of model prediction Equation 2.5, 2.6 and data in 2 stages: Second test-stage considering also mRNA half-lives for kidney.** Shown are scatter plots of predicted and measured relative amplitudes and phases. The blue lines indicate the theoretical prediction according to model with constant half-life Equation 2.1. I allow for a divergence of 20% of value range from the theoretical prediction, *i.e.* 20% of 24 hours for phases and 20% of 1 for relative amplitudes, indicated by the shaded area. mRNA abundances outside of this area I consider as "cannot be explained by transcriptional activity".

2.4 Quantifying the Extent of Rhythmic PTR

In the following section I will set out to reliably quantify the proportion of transcripts with a rhythmic PTR. Again, I will compare model prediction and measurement for transcript abundance, however this time I shall include all available information. This has the disadvantage that not every transcript can be tested or the test becomes very stringent and will only detect extreme cases.

2.4.1 PA-test – A Statistical Test with Half-Life

A good approach is to construct a statistical test which not only gives a yes-or-no-answer but also provides a p-value, a measure for the unlikeliness of the null-hypothesis, that is in this case "Oscillation in a transcript stems only from its rhythmic production".

To construct a statistical test based on the model results we require information about the uncertainties for all measured quantities. The uncertainties of phase and amplitude of transcript abundance and transcriptional activity are obtained by the residuals of the fit, see derivation in Appendix C.1.3. To estimate the errors of transcript half-life measurements, we would require more than one measurement per transcript, assume a Gaussian distribution and obtain its standard derivation. But only a small fraction of half-lives is measured in both Friedel *et al.* [212] and Schwanhäusser *et al.* [213]. This restricts the transcripts to which the test can be applied to only 23% of all circadian transcripts in kidney and 24% in liver. In Appendix C.2 I outline how I retrieve means and standard deviation for half-lives from Friedel *et al.* [212] and Schwanhäusser *et al.* [213].

The null hypothesis of the test is that an observed oscillation of a transcript stems only from the oscillation in its transcriptional activity. The model from Section 2.1 predicts a transcript abundance with only rhythmic transcriptional activity. If model and data do not agree within the assumed measurement errors, we can reject the null hypotheses.

In mathematical terms an oscillation is represented by a vector consisting of its cosine and sine fractions. Their uncertainties, depicted as ellipses, are defined by their two-dimensional covariance matrices, see Appendix C.1.3. To derive the covariance matrix in the model prediction, a quantity derived from transcriptional activity and half-life, we need to apply error propagation as outlined in Appendix A.2.1.

The test compares the two dimensional vectors describing transcript abundance derived from model prediction $\vec{a}_{\text{model}}$ and the fit of the time series $\vec{a}_{\text{measure}}$. I assume that both quantities are characterized by a two-dimensional Gaussian-distribution with means μ_{measure}, μ_{model} and covariance matrices Σ_{measure}, Σ_{model}, derived with error propagation (Appendix A.2.1). Then the mean difference between model prediction and measurement $\Delta = \mu_{\text{model}} - \mu_{\text{measure}}$ describes the deviation of the measurement from the model prediction. Its distribution can be approximated as a two-dimensional Gaussian distribution $N(\Delta, \Sigma)$ with covariance matrix $\Sigma = \Sigma_{\text{model}} + \Sigma_{\text{measure}}$.

To test if model prediction and measurement have the same mean leads us to a multivariate equivalent of a Student's t-test. This multivariate test statistic is characterized by Hotelling's T-squared distribution T^2:

$$T^2 = n\Delta^T \Sigma \Delta, \tag{2.7}$$

where n is the number of observations. It can be shown, that for a large number of observations T^2 follows a Chi-square distribution with, in this case, 2 degrees of freedom [215]:

$$T^2 \sim \chi^2_2. \tag{2.8}$$

The estimation of covariance matrices from the fit of time series and estimation of p-values according to the chi-squared statistic were implemented by Paul Thaben as the function pa.test in the R package "patest".

Figure 2.6A shows an illustration of the test. Note that we predict from transcript abundance together with the half-life the transcriptional activity. This prediction is then compared to the actually measured transcriptional activity. For each gene a different test will be performed, each with its own oscillation properties, half-lives and measurement errors to provide the best possible accuracy.

Another major advantage of this test is to avoid detecting a disagreement between model prediction and measurement solely based on the binary decision, either rhythmic or arrhythmic as has been done in the previous section. In this mind set, an arrhythmic transcriptional activity alone is not able to produce rhythmic transcript abundances. Camk2b, an example from liver, Figure 2.6B, proves the opposite. It has an arrhythmic transcriptional activity, but a rhythmic mRNA abundance. However, the χ^2 test detects no significant difference between measured and predicted transcriptional activity due to the large noise in transcriptional activity. This proves arrhythmic transcriptional activity is able to produce rhythmic mRNA abundances.

In the following I will use the χ^2 test to quantify the proportion of transcripts under rhythmic post-transcriptional control.

To apply the test, we need an estimate of the transcript's half-life and its variance as mentioned earlier. Furthermore, I only investigated genes which are rhythmic in their abundance and/or their transcriptional activity (false discovery rate below 0.25 and relative amplitude larger than 0.1). These restrictions result in 899 genes in kidney and 1023 genes in liver from the previously introduced data sets.

I estimated phase and amplitude of each time series by fitting a sine-function using the R-package Harmonic.Regression implemented by Pål Westermark. The package further provides the sum-squared-residuals (ssr) of the fit. With information on phase, amplitude and their uncertainties for transcript abundance and transcriptional activity, together with the mean and standard deviation of half-life measurements of Friedel *et al.* [212] and Schwanhäusser *et al.* [213] I can now apply the statistical test. It reveals that 158 out of 899 (18%) genes in kidney and 349 out of 1023 (34%) in liver have a rhythmic PTR (FDR ≤ 0.25), Figure 2.6D.

Comparing transcripts in both organs reveals a very small overlap. 393 genes were tested in both organs, but only 44 of these show a positive test result for rhythmic PTR (FDR $\leq$ 0.25). This suggests, that not only the circadian abundance of many transcripts is organ specific but also their rhythmic PTR. I will discuss the test result in more detail in Section 2.7. Figure 2.6C shows the Cold-inducible RNA binding protein (Cirbp) which ranked low according their q-values in both organs. This figure also illustrates what a significant test-result means in the vector representation. Cirbp is also already experimentally shown to be under rhythmic post-transcriptional control [170]. Interestingly, Cirbp itself is known to control circadian gene expression by affecting PTR in a rhythmic manner [169].

2.4.2 Test without Half-Life

The previous test results made use of measured half-lives. However, half-life measurements are only available for a small fraction of transcripts. Is there a way to test every transcript for rhythmic PTR regardless of its half-life? Ignorance of the half-life needs to be compensated by taking into account every possible half-life. This will weaken the power of a possible test significantly, but may still yield more insight on the data structure and provide potential for detection of PTR.

The overall idea for this test follows the geometrical consideration from the previous Section 2.3 but additionally includes the uncertainties of each oscillation derived from the sum-squared residuals of the sine-fit.

Any oscillation in the transcriptional activity is represented by a two-dimensional vector with an error ellipse. The model from Section 2.1 dictates that a rhythm in the transcript abundance appears on a point on the edge of a semi-circle if the rhythm is only influenced by a rhythmic transcriptional activity. The exact point is defined by the half-life. The semicircle spans between transcriptional activity and the origin, see Figure 2.6A. These two points correspond to very short transcript half-life, no phase shift and no amplitude reduction on one hand and very long half-life, maximal possible phase shift of 6 hours and complete loss of oscillation on the other hand, see also Figure 2.4A. The line of the semicircle describes only one possible oscillation derived from the mean of the distribution of the transcriptional activity. Taking into account the uncertainty or measurement error of transcriptional activity means that an error ellipse of transcriptional activity is spanned from each point on the semicircle. This yields an area which contains all possible transcript abundances which could be produced by the specific transcriptional activity distribution, see Figure 2.6A. If an oscillation of a transcript abundance cannot be explained by the transcriptional activity then the transcript abundance and its error ellipse will lie outside of this semicircle.

Note that, in this case, the transcriptional activity is used to predict possible transcript abundance. This approach is the other way around than the previous test, where transcript abundance was used to predict the transcriptional activity, see Figure 2.6A.

To apply the test I follow a similar procedure as previously. Only genes which have a circadian oscillation (FDR ≤ 0.25 and relative amplitude larger than 0.1) in either their transcriptional activity and/or their transcript abundance are tested. For each of these genes I fit a sine curve to transcriptional activity and transcript abundance using the package Harmonic.Regression, which also provides the sum-squared residuals to construct the error ellipses. For each transcriptional activity I construct a semicircle-area as described above and check if there is an overlap with the error ellipse of its associated transcript abundance. In this way, only 52 genes in liver and 11 genes in kidney were tested to have a rhythmic PTR.

The size of error ellipses influences the outcome of this test significantly, smaller error ellipses would decrease a possible overlap. Smaller error ellipses can be achieved if the oscillation, a dot in the vector plane, is measured more often and hence the confidence of the measurement is increased. In Atger *et al.* [216] transcriptome of mouse liver was measured by RNA sequencing with a high sampling frequency, every 2 hours over the course of 4 days, in total 48 samples. An estimation of the transcriptional activity can be achieved from the intronic reads of RNA sequencing as has been done for the kidney data set.

Equipped with this data I receive 8109 out of 12351 (66%) circadianly expressed genes, either in transcriptional activity or transcript abundance. Here, already the higher confidence of oscillation is visible in the higher percentage of circadian genes. This can be seen even better when we apply the test. Here, 69 genes instead of 52 genes in mouse liver were tested to have a rhythmic PTR. However, this result is still far lower than the results of the test which uses half-lives.

In conclusion, I described two tests which compare model prediction with measurement. The first test, a statistical test taking into account the transcript half-lives was used on a representative subset of the circadian transcriptome liver and kidney and found that 18% of the circadian transcriptome in kidney and 34% in liver are under rhythmic PTR. The second test does not require the transcript half-life. However, with this test only very few genes can be detected. The test could be used to find genes for which a rhythmic PTR has a strong influence. Compared to the χ^2-test the second test (without half-life) revealed a few genes which were not detected by the first test due to a missing measured half-life: 10 genes in kidney and 48 genes in liver. Accordingly, 1 and 4 genes for kidney and liver, respectively, were detected by both tests.

For convenience of the reader, the whole set of genes positively tested for rhythmic PTR is published in Appendix F.3. Can you find your favorite gene?

Additionally, the numbers of how many circadian genes and which genes were tested from Sections 2.2, 2.3 and this section are summarized in a table in Appendix F.2.

Having quantified the amount of genes under rhythmic PTR the following questions emerge: Which post-transcriptional processes can explain the discrepancies between model prediction and measurement? Is rhythmicity in any step of PTR sufficient to explain the findings? To answer these questions I will introduce in the following a second model which describes not only rhythmic production, but also rhythmic degradation. This will be the starting point to investigate other post-transcriptional processes.

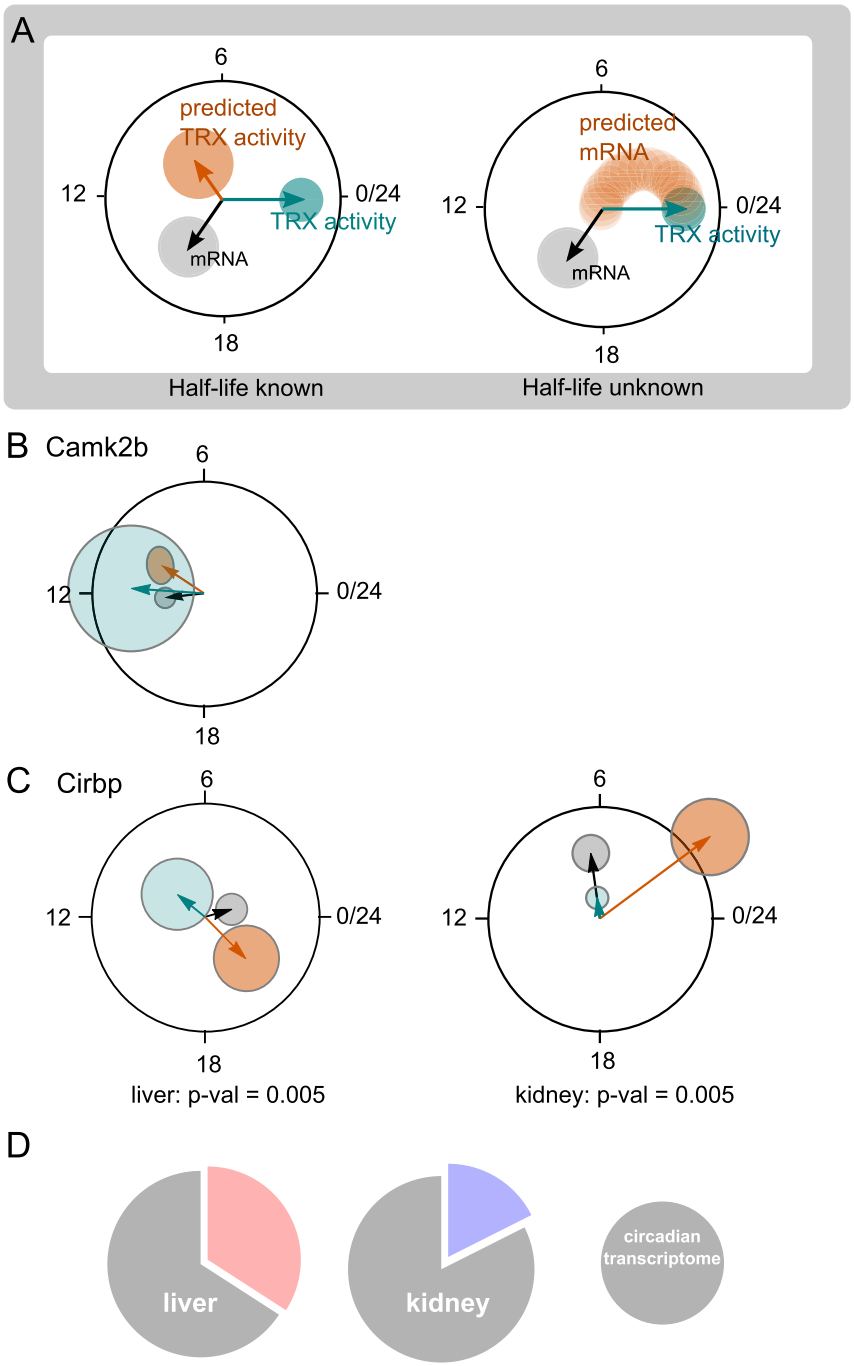


Figure 2.6: **How many genes are under rhythmic PTR? A** Illustration of two tests, one with (left), the other without considering the half-life. Blue: transcriptional activity, production rate in the model. Black: mRNA abundance, product x in the model. Orange: predicted quantity. For each oscillation the vector and its corresponding ellipse stands for the oscillation and its uncertainty. In the left test transcriptional activity is predicted from mRNA abundance (orange) and compared with the measured transcriptional activity. For the other test the mRNA abundance is predicted from transcriptional activity. **B** Test result for Camk2b from statistical test using half-lives. Transcriptional activity is arrhythmic, mRNA abundance is rhythmic. However, prediction of transcriptional activity (orange) does not differ from measurement. **C** Test results from statistical test using half-lives for Cirbp in liver and kidney. **D** Summary of statistical test: From a representative test set 34% genes in liver and 18% in kidney are positively tested for PTR (FDR ≤ 0.25).

2.5 Rhythmic Post-Transcriptional Degradation

More than 15% and up to one third of the circadian transcriptome cannot stem from rhythmic transcription alone, see previous section. The search for additional rhythmic processes influencing mRNA abundance leads the focus to PTR. To investigate rhythmic PTR I will again use an ODE model similar to the one presented in Section 2.1. Instead of rhythmic production and constant degradation, I will now describe an mRNA which is rhythmically produced and rhythmically degraded. I will use this model further not only to examine rhythmic degradation but also to answer the question if other rhythmic post-transcriptional processes such as rhythmic splicing or rhythmic nuclear export can explain the findings.

2.5.1 An ODE Model: Production and Degradation with Oscillating Rates

An mRNA x is rhythmically produced and rhythmically degraded with circadian rates (see Figure 2.7A):

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \mathrm{prod}(t) - \mathrm{deg}(t)x, \\ \mathrm{prod}(t) &= k\left(1 + A_{\mathrm{prod}} \cos\left(\omega t - \phi_{\mathrm{prod}}\right)\right), \\ \mathrm{deg}(t) &= k\left(1 + A_{\mathrm{deg}} \cos\left(\omega t - \phi_{\mathrm{deg}}\right)\right). \end{aligned} \tag{2.9}$$

The time dependent rates are described by cosine functions with angular frequency ω, means k and γ, relative amplitudes A_{prod} and A_{deg}, phases ϕ_{prod} and ϕ_{deg} for production and degradation rate, respectively. The angular frequency is set to $\omega = \frac{2\pi}{24}$ h^{-}1 to reflect circadian dynamics. The mean degradation rate is connected to the half-life as described before in Equation 2.2.

The solution $x(t)$ of Equation 2.9 cannot be calculated analytically. I therefore use a Fourier expansion to find an approximation for the solution. The derivation is described in Appendix A.1.2. Furthermore, I verified, that this approach is suitable for this problem. Specifically, I verified that Fourier terms in the approximation vanish for higher order and I quantified the error made by the approximation by comparing it to the numerical solution of the model Equation 2.9. This inspection of the approximated solution is published in Appendix A.1.3 and A.1.4.

The solution can be understood intuitively as vector calculation. Rhythms in production and degradation rate characterized by phase and amplitude are represented by vectors in the 24-hour plane as described in Section 1.3.1. The vector of mRNA abundance, the solution of the model Equation 2.9, is then determined by first taking the vector difference of production and degradation rate. We termed this vector difference "production-degradation-vector", see Figure 2.7B. The production-degradation-vector is further shifted and decreased in length by very similar half-life-dependent factors to the ones already received from the model of constant degradation (Equation 2.6, 2.5). Hence, it follows that this solution in the limes of constant degradation falls together with the solution of the previous model describing constant degradation.

In mathematical terms the vector describing the mRNA abundance is characterized by its

direction or mRNA phase ϕ_x and the vector length, the oscillation's relative amplitude A_x:

$$\phi_x = \underbrace{\arg\left(A_{\text{prod}}e^{i\varphi_{\text{prod}}} - A_{\text{deg}}e^{i\varphi_{\text{deg}}}\right)}_{\text{production-degradation vector phase}} + \underbrace{\arctan\left(\frac{\omega}{\gamma}\right)}_{\substack{\text{phase shift determined} \\ \text{by mean half-life}}} \tag{2.10}$$

$$A_x = \underbrace{\left|A_{\text{prod}}e^{i\varphi_{\text{prod}}} - A_{\text{deg}}e^{i\varphi_{\text{deg}}}\right|}_{\substack{\text{Combined relative amplitude of} \\ \text{production and degradation}}} \times \underbrace{\frac{\gamma}{\sqrt{\gamma^2+\omega^2} - \frac{1}{\sqrt{\gamma^2+\omega^2}}C}}_{\substack{\text{amplitude reduction} \\ \text{determined by mean half-life}}} \tag{2.11}$$

$C = A_{\text{prod}}A_{\text{deg}}\gamma/2 \times [\omega\sin(\varphi_{\text{deg}} - \varphi_{\text{prod}}) + \gamma\cos(\varphi_{\text{deg}} - \varphi_{\text{prod}})]$ is a factor which is small for small relative amplitudes of the rates.

What are the consequences of this model solution for the abundance of a transcript? The vector representation provides an intuitive understanding of the influence of rhythmic degradation. This might be helpful for researchers who do not have an advanced education in mathematics. Furthermore, the structure of the solution facilitates a systematic understanding. It allows for a separate discussion of the influence of the mRNA's mean half-life from the influence of the rate's oscillatory properties, namely relative amplitude and phase.

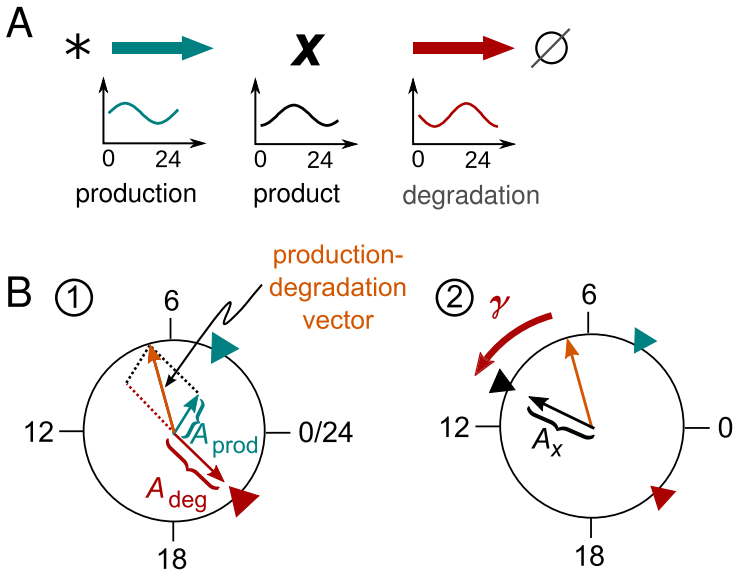


Figure 2.7: **A model describing rhythmic production and rhythmic degradation and its solution.** **A** An mRNA x is rhythmically produced and degraded. Both rates are modeled as cosine curves. **B** The solution of the model. Rates are represented as vectors, blue: production, red: degradation. To retrieve the mRNA abundance x, first, the production degradation vector is calculated as a vector difference of production and degradation rate, second, this vector is rotated and reduced in length by a factor which depends mainly on the transcript half-life.

2.5.2 Rhythmic Degradation Explains All Observed Phases and Amplitudes.

Phase and amplitude of a transcript is determined by the production-degradation-vector and therefore by the production and degradation rates' properties.

The rates' amplitudes can add up and form an amplitude boost if the phase relation between both rates is appropriate. The maximal amplitude is reached when production and degradation are in antiphase. A reduction of oscillation occurs if both rates oscillate in phase. The reduction is strongest for similar oscillation strengths. When production and degradation rate have the same oscillation strength and have exactly the same phase the length of the production-degradation vector is zero and there is no oscillation in the mRNA abundance, see Fig 2.8A.

The phase of production-degradation-vector, and hence the transcript phase, is dictated by the phase of the dominant rate, *i.e.* the rate with the strongest oscillation. A transcript peaking 6 to 12 hours before the production peak can only be realized when the degradation rate dominates, Figure 2.8B. Such phase relationships have been observed in both kidney and liver, see also Figure 2.4B. Furthermore, if a production and degradation rate oscillate with similar strength, *i.e.* the ratio of relative amplitudes is close to 1, the transcript's amplitude is low, but - at the same time - the biomolecule's phase is highly sensitive to changes in the relative amplitude in production or degradation. This is reflected by the sensitivity coefficient, Figure 2.8C and Appendix A.2.2.

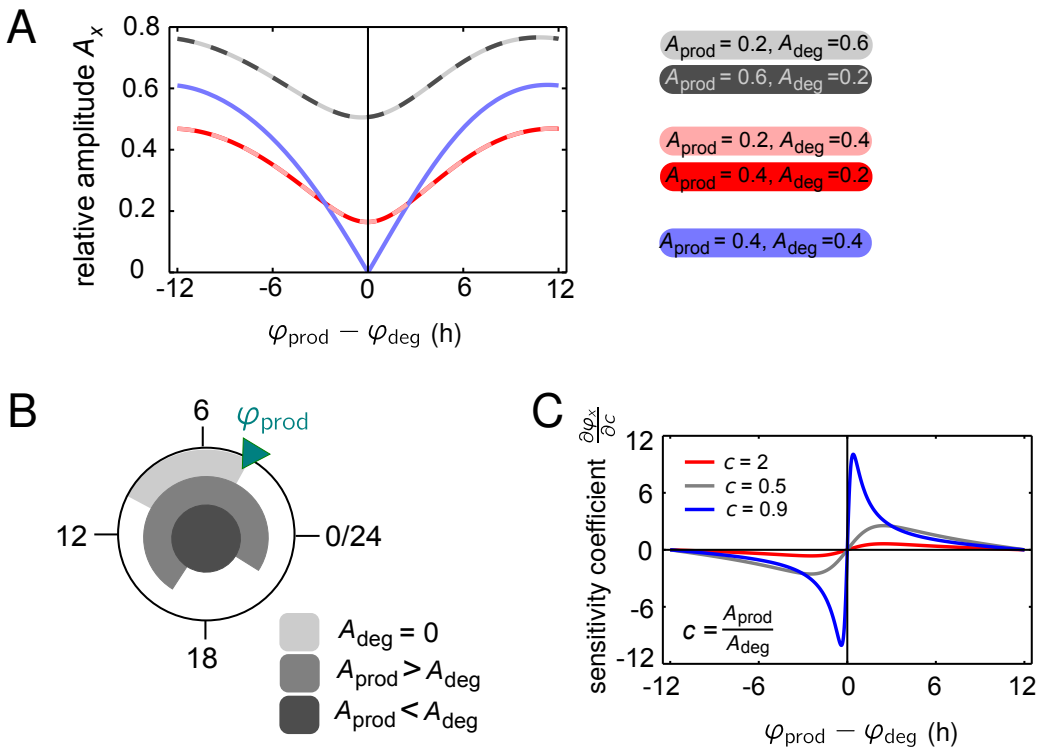


Figure 2.8: **Conclusions drawn from model solution: rhythmicity properties of the rates.** **A** Production and degradation can cancel or boost each other depending on phase and amplitude relationship. Shown is the model solution for different parameter sets of degradation and production rate. Half-life $t_{1/2} = 2$ h. **B** Possible phase ranges of mRNA abundance phase for different relations of production and degradation amplitudes. An mRNA which peaks 12 to 6 hours before its transcriptional activity must be influenced by a dominating oscillation in the degradation. **C** The sensitivity coefficient characterizes the change of mRNA phase under slightly different rate parameters. There is a high sensitivity, *i.e.* the mRNA phase is highly susceptible for changes in rate parameters, if production and degradation oscillate with a similar strength and phases are similar (see blue curve). A complete derivation of the coefficient is shown in Appendix A.2.2.

The production-degradation vector is further shifted and decreased in length. This shifting and shorterning is dictated by factors depending mainly on the transcript's half-life. These are very similar to the factors, which already describe the phase shift and amplitude decrease in the model of constant degradation, Equation 2.6, 2.5. Consequently, the same conclusions are also true for rhythmic degradation. For long mRNA half-life the oscillation of mRNA abundance vanishes accompanied by a phase shift of up to 6 hours. This was also verified through numerical simulations of the model, Figure 2.9A.

However, an additional rhythm in degradation can lead to an amplitude boost. Because of this additional amplitude gain, transcripts with longer half-lives can oscillate with functional rhythms. Let us assume only rhythmic mRNA abundances with amplitudes larger than 0.1 are biologically functional. Then, the half-life of transcript influenced only by rhythmic production can take values of up to 26.5 hours and still be biologically functional. However, transcripts under the influence of rhythmic production and rhythmic degradation can have a half-life twice as large, up to 53 hours, due to the amplitude boost. These values depend strongly on the rates' amplitudes and phase differences, Figure 2.9B.

In conclusion, the model solution tells us that the phase range and amplitude range of transcripts under the control of rhythmic production and rhythmic degradation spans between the extreme values, all possible values between these can be reached. Consequently, rhythmic degradation is able to explain the discrepancy between transcriptional activity and transcript abundance found in mouse liver and kidney. What about other PTR steps? Can a rhythm in any PTR explain the discrepancies? Or is there something special in rhythmic degradation? This will be looked at in the next main section. However, before we proceed I briefly present how possible degradation rates can be predicted using the ODE model. This is useful since in the data we tackle degradation is the unknown variable.

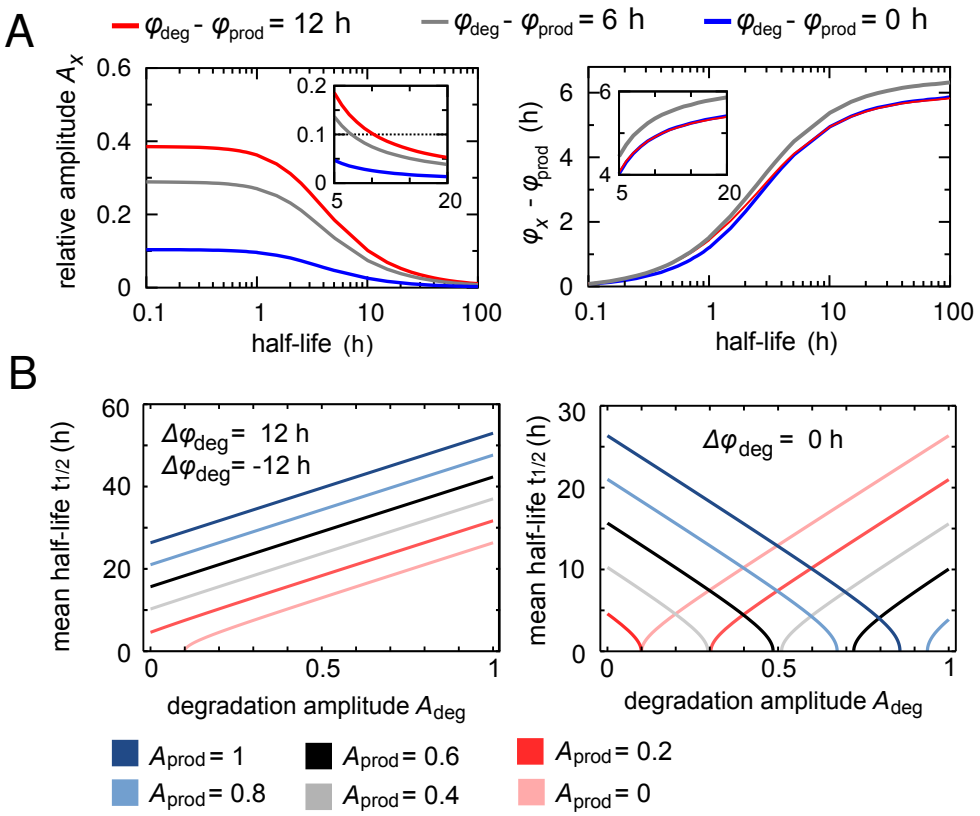


Figure 2.9: **Conclusions drawn from model solution: influence of half-lives. A** Phase shift and amplitude reduction for long half-lives if mRNA also is influenced by rhythmic production and rhythmic degradation. Shown are numerical simulations. Dependence of relative amplitudes and phases on the half-life. Parameter: Aprod $= 0.25$, $A_{\mathrm{deg}} = 0.15$, $k = 1\ \mathrm{h}^{-1}$, $t_{1/2} = 2$ h. **B** The relative amplitude A_x vanishes for long half-lives. Shown are maximal mRNA half-lives required to generate a relative amplitude $A_x > 0.1$ for different rate parameters.

2.5.3 Predicting Possible Degradation Rates

The model Equation 2.9 uses given production and degradation rates to predict the transcript abundance. The data I investigate contains information of production rate and the transcript abundance. The degradation rate is not measured.

However, I can use the approximated solution of the model to predict an unmeasured degradation rate by transforming Equations 2.10, 2.11.

The vector annotation provides an intuitive understanding of this transformation. Phase and amplitude of an unknown degradation can be estimated by the reverse calculation of the model solution. The vector of the now known molecule abundance x is rotated backwards and increased in length by a half-life dependent factor. From the resulting production-degradation vector the production-rate-vector is subtracted. The resulting vector is the degradation rate-vector in opposite direction, see Figure 2.10.

In mathematical terms this reads:

$$\phi_{\mathrm{deg}} \approx \arg\left(A_{\mathrm{prod}} e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{pd}} e^{i\phi_{\mathrm{pd}}}\right), \tag{2.12}$$

$$A_{\mathrm{deg}} \approx \left|A_{\mathrm{prod}} e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{pd}} e^{i\phi_{\mathrm{pd}}}\right| \tag{2.13}$$

where the relative amplitude A_{pd} and phase φ_{pd} of the production-degradation vector are calculated using the half-life and the phase and relative amplitude of the molecule, yielding $A_{\mathrm{pd}} = A_x\sqrt{\gamma^2+\omega^2}/\gamma$ and $\phi_{\mathrm{pd}} = \varphi_x - \arctan(\omega/\gamma)$.

In order to verify this approach, I estimated the degradation rates from numerical simulations with different parameter sets for rhythmic production and degradation rates. I found there is a reasonably good agreement between degradation rate used for the numerical simulation and estimated degradation rate. Especially for low relative amplitudes in the rates, the estimation of the degradation phase deviates by up to 2 hours (12% of 24 hours) from the numerical simulation and the degradation amplitude deviates by up to 0.2 (20% of the maximal relative amplitude), see Appendix A.1.4. This is useful for an application since most transcript abundances have small relative amplitudes, which would translate into even smaller relative enzyme abundances and possibly activity rates and hence production and degradation rates might often oscillate with a small amplitude.

2.6 Rhythms in Other Stages of PTR

The previous model describes one post-transcriptional process, degradation, to be rhythmic. Is it possible to use the same model to investigate any other post-transcriptional process? Can other processes besides degradation explain discrepancies between transcriptional regulation and transcript abundance? Or, in other words, which post-transcriptional process can result in a phase shift of more than 12 hours between transcriptional activity and mRNA abundance and which processes are able to increase amplitudes? These are findings which are seen in the data and cannot be explained simply by transcriptional activity. In the following I will examine other post-transcriptional processes to investigate if any of these can explain observed phase and amplitude relationships. Since PTR is complex involving different steps acting on

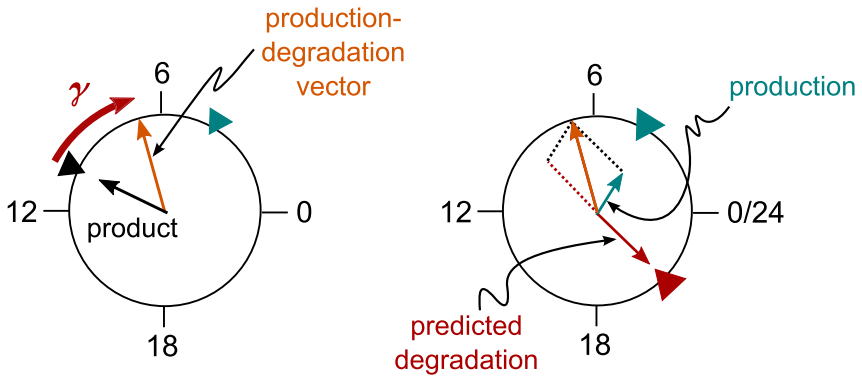


Figure 2.10: **Illustration of predicting an unknown degradation rate with given mRNA abundance and transcriptional activity (=production).** It is basically the reverse calculation made in Figure 2.7B.

different stages during an mRNA life-time, my goal is to develop a system - making use of the preiovous results - which allows the classification of any arbitrary, even yet unknown, rhythmic post-transcriptional process. I will then use this system to investigate mRNA splicing, besides mRNA degradation another post-transcriptional process known to be rhythmic.

2.6.1 Three Questions Guide the Examination of Other PTR

To examine any post-transcriptional process using the previous results one has to decide if a process is reflected by the available data, and if and how this process is represented by the model of rhythmic production and rhythmic degradation. This decision can be made by answering three questions: Does the post-transcriptional process of interest change any of the measured quantities? Does this process destroy a pre- or mature mRNA or does it reversibly change its state? And finally, does this process affect pre-mRNA or mature mRNA?

The first question shows if a post-transcriptional process is captured by the available data. If this is not the case the process naturally cannot explain observed discrepancies. A counterexample would be mRNA localization. This post-transcriptional process cannot be seen in the RNA sequencing result because RNA-seq captures all cytosolic RNA independent from their position in the cell. Consequently, a rhythmic RNA localization cannot explain any observed phase or amplitude relations.

The second question shows if the transformation of the RNA is irreversible as in splicing or in mRNA degradation. A reversible post-transcriptional modification, m^6 mRNA methylation for example, cannot be reflected by the model. A rhythmic reversible reaction changes the structure of the model and consequently its dynamics so drastically that it cannot be translated to the already studied dynamics. However, first numerical results reveal that even with a reversible, rhythmic post-transcriptional modification rhythms with any phase in mRNA abundance could be realized (not shown). But extensive investigation of reversible post-transcriptional modifications might not be appropriate for the available data, since to my knowledge no reversible modification can be captured with the used experimental methods. Also the theoretical example

of m^6 mRNA methylation would be filtered out already by the first question.

After ensuring that the post-transcriptional process of interest is captured by the data and that this process corresponds to the overall dynamics of the analyzed model the last question reveals if the model must be extended in order to fully describe the rhythmic post-transcriptional process.

There is no need for an extension if the post-transcriptional process of interest removes mature mRNA with no return. Besides mRNA degradation, this could be realized, for example, by rhythmic mRNA export out of the cell [217]. Every conclusion drawn from the previous analysis is also true for these processes and they serve for an explanation of the observed phase and amplitude discrepancies between transcriptional activity and mRNA abundance.

However, if the rhythmic post-transcriptional process of interest acts on pre-mRNA, it is necessary to expand the model by one additional species. In this expanded model, the pre-RNA is rhythmically produced by transcriptional activity and rhythmically transformed to RNA or in other words rhythmically "destroyed" as pre-RNA. The rhythmic transformation serves as rhythmic production of the biomolecule mature RNA. In mathematical terms this reads

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \mathrm{prod}_x(t) - \mathrm{deg}_x(t)x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \mathrm{deg}_x(t)x - \mathrm{deg}_y y, \\ \mathrm{prod}_x(t) &= k\left(1 + A_{\mathrm{prod}}\cos\left(\omega t - \phi_{\mathrm{prod}}\right)\right), \\ \mathrm{deg}_x(t) &= L\left(1 + A_{\mathrm{deg}}\cos\left(\omega t - \phi_{\mathrm{deg}}\right)\right), \\ \mathrm{deg}_y &= \gamma. \end{aligned} \tag{2.14}$$

Here, $\mathrm{prod}_x(t)$ and $\mathrm{deg}_x(t)$ are the time dependent production and degradation rate of pre-RNA x with relative amplitudes A_{prod}, A_{deg} and phases ϕ_{prod}, ϕ_{deg}, respectively. L is the mean rate of the rhythmic post-transcriptional process. The mature mRNA y is constantly degraded with rate γ. Although constituting a new system the dynamics of this ODE system can be broken down to the already known solutions stated in the previous section. The first equation describes a biomolecule being rhythmically produced and degraded, the second can be interpreted as RNA with rhythmic production and constant degradation. The phase difference and amplitude relation between measured transcriptional activity and mRNA abundance is then stated by the phase and amplitude relation between rhythmic production of x and the abundance of y. It is given by combining solutions from the two models. The rhythmic post-transcriptional process in this formulation acts as rhythmic degradation of the pre-RNA. The model of rhythmic production and rhythmic degradation dictates that production and degradation rate or rhythmic post-transcriptional process form the production-degradation vector which is shifted and reduced by a factor depending on the mean rate of the rhythmic post-transcriptional process. The resulting vector is then further shifted and reduced in length by a factor depending on the mRNA half-life, as described by the model of constant degradation, see Sec. 2.1. Consequently, an unknown rhythmic post-transcriptional processing of pre-RNA can also result in any phase and amplitude relation between transcriptional activity and mRNA abundance. However, a possible amplitude boost, arising from an advantageous relation between rhythmic rates of transcriptional

activity and post-transcriptional process, is subject to a stronger decrease in amplitude due to more processing steps, *i.e.* rhythmic post-transcriptional processing and mRNA degradation, between generation of the oscillation and measurement of mRNA abundance.

In summary, with a systematic analysis guided by three questions, I systematically examine the influence of a rhythmic post-transcriptional process on available data. I first ensure that any particular rhythmic post-transcriptional process is captured by the data and can be described by the model of rhythmic production and rhythmic degradation. I then use the previously achieved results to deduce if this rhythmic process can explain every observed phase and amplitude in mRNA abundance.

I will discuss this approach using the process of splicing. Rhythmic splicing of certain transcripts occurs due to the rhythmic abundance of splicing-(co-)factors, availability of splicing sites due to changed conditions [218] or possibly due to variation of transcriptional speed which changes the availability and consequently the cooperativity between splice sites. A single gene can code for many proteins. This is realized by so-called alternative splicing, *i.e.* including or excluding certain exons due to different splicing conditions. In mouse liver it is estimated that 20% of expressed genes contain a circadian exon [186].

2.6.2 Systematic Analysis of Rhythmic Splicing

Can rhythmic splicing result in any phase and amplitude relation between transcriptional activity and mRNA abundance? Let us follow the proposed system to answer this question. The data on pre-RNA or transcriptional activity represents unspliced RNA, RNA sequencing accounts for spliced RNA. Splicing is hence represented by the data. Furthermore, splicing transforms pre-RNA into mature RNA with no return, the overall dynamics of the model of rhythmic production and rhythmic degradation is valid. Splicing changes the level of pre-RNA. The process is therefore described by an 2-ODE-system, where a pre-RNA (x in model model 2.14) is rhythmically transcribed and rhythmically spliced, the rhythmic splicing serves as a rhythmic production of mature RNA (y in model 2.14). The phase and amplitude of an mRNA is then determined by the phase and amplitude relation of transcriptional activity and splicing. Since the mean splicing rate is fast (see Section 2.1), the mRNA abundance is mainly influenced by the transcript's half-life. To have any resulting phase relationship between pre-RNA and mature RNA, observations that we seek to explain, splicing should oscillate independently from transcriptional activity. However, in many cases splicing occurs co-transcriptionally [198] and is thus timely coupled to transcriptional activity. Rhythmic splicing that is not coupled to a rhythmic transcription would furthermore lead to accumulation of unspliced transcripts in the nucleus. However, RNA accumulation in the nucleus has, to date, been rarely observed and is rather a reaction to stress signals [219, 220, 221] than a common cellular program. Hence, rhythmic splicing resulting in only one transcript is not supported by existing knowledge on splicing and furthermore, would fail to explain a wide range of phase relationships owing to the coupling of transcription and splicing and hence similar phases in the rates.

Alternative splicing, however, produces 2 or more transcripts. The switching between one transcript variant to another due to alternative splicing is not coupled to the transcription itself. Transcription rate and splicing rate of a certain transcript variant can take any phase relationship and, hence, circadian alternative splicing together with circadian transcription can

produce any phase and amplitude in mRNA abundance.

In summary, many phase and amplitude relationships between transcriptional activity and mRNA abundance cannot be explained by rhythmic transcription alone. Consequently, mRNA abundance is further influenced by other rhythmic post-transcriptional processes. Only rhythmic processes which are reflected by the data and are not timely coupled to the transcription can explain the wide range of phases in mRNA abundance. Processing of pre-RNA, however, is often coupled to transcription, for example adding the 5'cap is completed before the end of transcription, most splicing occurs co-transcriptional [198] and the poly(A) tail is added once the transcription is terminated. Regulating nuclear export seems to be transcription independent and if being rhythmic, nuclear export could possibly explain observed phase relation ships. However, rhythmic nuclear export together with uncoupled rhythmic transcription would lead to RNA accumulation in the nucleus. As explained earlier, this does not seem to be a regular cellular program such as circadian gene expression. With that I conclude, up to the current state of knowledge, only three biological processes are able produce every relationship between transcription and mRNA abundance, that is rhythmic alternative splicing, rhythmic degradation and rhythmic mRNA export out of the cell.

2.7 Characterizing and Analyzing Rhythmic PTR in Liver and Kidney

About 18% transcripts of all circadian transcripts in kidney and 34% in liver are positively tested for rhythmic post-transcriptional control, see Section 2.4. With that the fraction of genes with rhythmic PTR is twice as large in liver compared to kidney. Only processes with an independent phase from transcription can explain these findings, see previous section. Furthermore, it is possible to predict the phase and amplitude of the rhythmic post-transcriptional process, see Section 2.5.3. All of these findings are based on mathematical modeling. This is of course only a hint towards actual evidence.

Is it possible to support these findings with additional evidence derived from a different source? Already, in our paper Lück *et al.* [179] we showed that genes positively tested for rhythmic PTR are also overrepresented in targets for CIRBP (Cold induced RNA binding protein), a protein which has a circadian abundance and destabilizes transcripts [169]. Furthermore, we showed that predicted degradation rates correlate with measured rhythmic poly(A) tail lengths [189]. Poly(A) tail length of a transcript is connected to its half-life. These two findings verify that RNAs we classify using our test as under rhythmic post-transcriptional control are also found experimentally to be under rhythmic post-transcriptional control.

Is there further evidence? Why are the fractions of circadian transcripts with rhythmic PTR so different between both organs? To answer these question I will try to find the "footprints" of rhythmic PTR, either in biologically motivated evidence or based on the shape of mRNA abundance time series.

2.7.1 Circadian Genes Have Longer UTRs in Liver than in Kidney

Any PTR is mediated by so-called trans-factors, which bind to their recognition sites in the un-translated regions (UTRs) of mRNA. It is interesting to see that genes circadianly expressed

in liver have slightly longer UTRs in liver than kidney, see Figure 2.11. This difference is not observed for all expressed genes in liver and kidney, see Figure 2.11. With longer UTRs and presumably more recognition sites transcripts are prone to a more complex PTR. This might explain the larger proportion of genes positively tested for rhythmic PTR in liver.

UTRs contain recognition sites for trans-factors, known examples are miRNA and RNA binding proteins. The first, miRNA, are, most likely, not able to influence circadian rhythms because timescales in miRNA-dependent regulation are too long [178], see also Section 1.2.1. Therefore, I will concentrate in my investigations on RNA-binding proteins (RBPs). RBPs bind to the mRNA and determine the mRNA's fate by controlling post-transcriptional processing on every step of an mRNA's life time. More than 1500 RBPs have been identified in humans [151], each usually targeting groups of mRNAs rather than single mRNAs. These groups of genes are then each co-regulated by one RBP. It has been argued that each group encodes functionally related proteins [222]. I will therefore investigate, which RBP are rhythmic in mouse liver and kidney. Furthermore, I will investigate if binding sites of these RBPs are enriched in those genes which are positively tested for rhythmic PTR. And finally, I will investigate if those genes group together regarding their gene function which would support that there is a co-regulation.

2.7.2 Rhythmic RBPs and Predicted Degradation Rates are Different in Both Organs

A rhythmic RBP activity is induced by three situations: First, a rhythmic RBP abundance; second, a constant abundance but rhythmic protein activity caused by post-translational modifications or third, the rhythmic appearance of RBP recognition motives on the transcript itself mediated by alternative splicing. It is estimated that 20 % of circadian genes contain a circadian exon [186]. Likewise, one can assume the regulatory region and with that availability of RBP binding sites to be affected by alternative splicing. However, to characterize alternative splicing of UTRs in circadian genes is far beyond the scope of this thesis. To my knowledge, there is no comprehensive investigation of RBP activity available. Rhythmic abundance of RBPs in mouse liver and kidney, however, can be estimated from their transcript abundance. I used three curated databases to find a comprehensive list of RBPs. I excluded those with non-circadian transcripts and transcripts which have only a small relative amplitude. Additionally, I excluded those, for which a long protein half-life was reported [213]. In this way I found 25 RBPs in liver and 15 RBPs in kidney whose transcript have a circadian abundance, which is an indicator for rhythmic protein abundance. Only 4 RBPs have rhythmic RNAs in both organs. A complete list, together with known gene functions, can be found in Appendix F.4. Phase distributions of the RBP transcript phases reveal that in both organs certain phases are more prominent, see Figure 2.12A, in kidney more RNA binding proteins are expressed during the first half of the day, in liver more transcripts are expressed in the second half of day. Although the phase of a transcript is not a reliable estimator for the protein phase, the accumulation of RBP transcript phases indicate that rhythmic PTR might be more pronounced at some specific times throughout the day.

This is supported by distributions of predicted degradation phases for genes with rhythmic PTR, see Figure 2.12A. Here, I also found an accumulation of degradation phases at some times of the day. Interestingly, these peaks in the distribution are almost antiphasic between

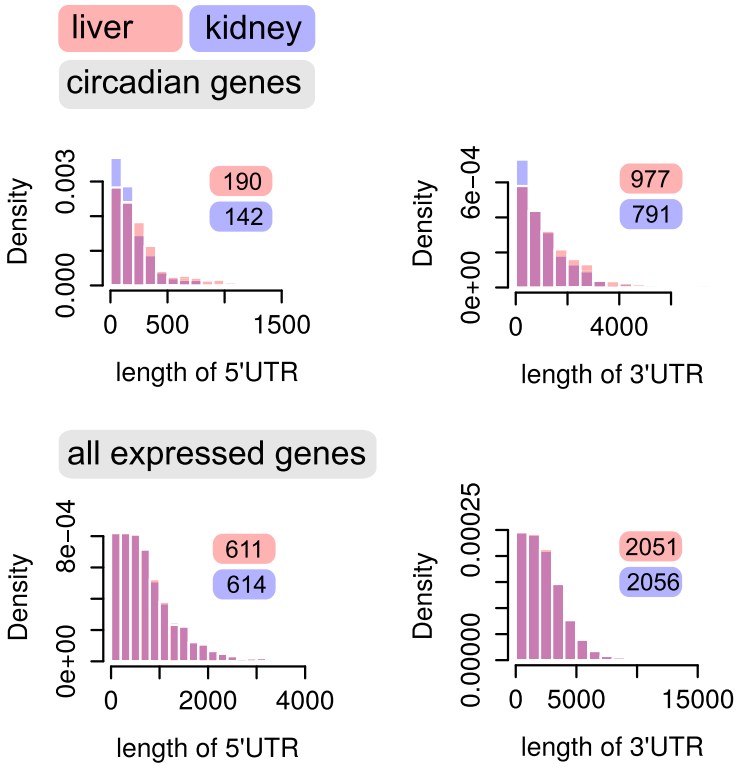


Figure 2.11: **UTRs of circadian genes are longer in liver than in kidney.** Shown are histograms of 5' and 3' UTRs from circadianly expressed genes or all expressed genes in liver and kidney. In order to allow for comparison, the density is shown. Numbers at each histogram indicate the median of distributions. Transparent red: liver, transparent blue: kidney, overlap of both appears violet.

both organs. Different phases in RBP transcripts and predicted degradation rates indicate that rhythmic post-transcriptional is highly organ specific, not only that there are different RBPs but also they act at different times during the day. Furthermore, PTR seems to cluster at specific times in both organs.

2.7.3 RBP Binding Sites and Gene Functions

Are genes with rhythmic PTR enriched in targets of specific RBPs? To investigate this I searched for RNA binding sites in the regulatory regions, 5'untranscribed region (UTR) and 3'UTR, of transcripts with rhythmic PTR. To test for enrichment I used a Fisher's exact test and tested an enrichment of bindings sites against binding sites in all circadian transcripts. With this I found that 4 RBPs in kidney and 70 RBPs in liver have significantly more targets in the group of transcripts with PTR than in all circadian transcripts. If these RBP truly mediate rhythmic PTR they would act at specific times of the day, that is to say, their action itself is circadian. Hence the predicted degradation phases of the RBPs' targets would accumulate at those times. However, the degradation phase distribution of the target genes for each RBP shows no pronounced accumulation of phases for any of the RBP (not shown). Instead, each distribution is similar to the distribution of all degradation phases.

It seems that looking at RBP binding sites is not an appropriate way to search for hints on rhythmic PTR. Most of the binding sites found are, most likely, not functional which results in many false-positives. This is supported by the fact that a larger number of significant RBPs have been found in liver. In liver rhythmic transcripts have much longer UTRs, which contain more RBP binding sites.

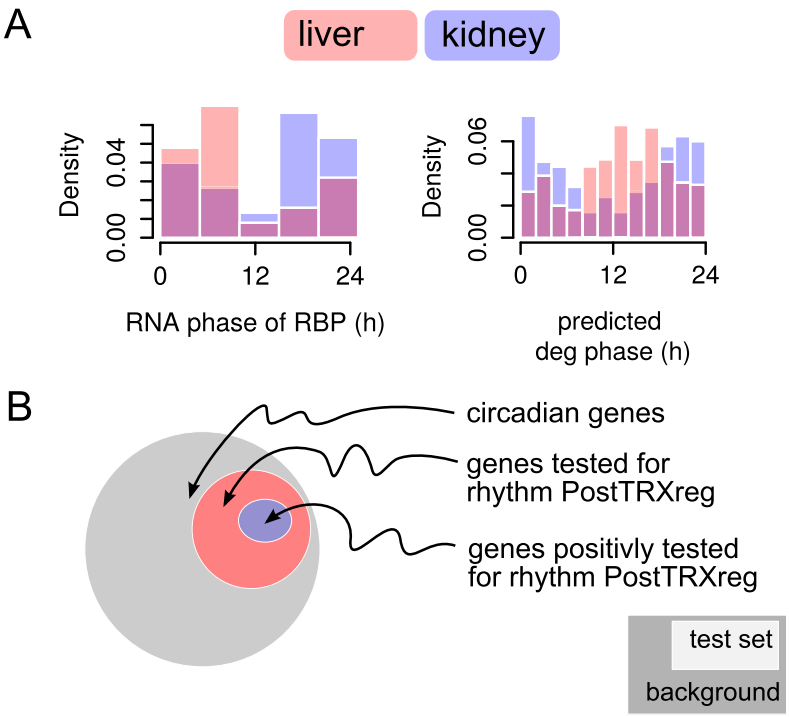
The search for RBP binding sites ended in a dead end due to unspecific binding sites. Although I could not pinpoint single RBPs which regulate groups of transcripts these groups might exist, since one RBP targets several RNA. These genes are possibly connected through their gene function [222].

Are genes with similar degradation phases and hence a possible co-regulation enriched for certain gene functions? To test for enrichment in gene functions I used the DAVID Function Annotation Tool [223, 224]. This tool uses a diverse range of annotation databases to identify gene functions. It then tests for enrichments of gene functions which are related in their biological function. These genes form gene function clusters. To test for enrichment the tool uses a Fisher's exact test where the background can be provided by the user. The tool then provides an enrichment score for each gene function cluster.

To use the DAVID Function Annotation Tool I carefully chose my background against which genes I wanted to test for enrichment. The genes I tested for rhythmic PTR had to fulfill certain requirements, namely rhythmicity in either transcript abundance or transcriptional activity and sufficient certainty about the half-life. These requirements limited the set of genes substantially. Hence, I first searched for enrichment in all genes tested for rhythmic PTR against all circadian genes. It turns out that this gene-set is highly enriched, *i.e.* high enrichment scores, for protein maintenance in a broad sense, see Figure 2.12B. I now tested the genes with rhythmic post-transcriptional for a gene function enrichment against all tested genes, see Figure 2.12B. I also found gene function clusters which relate to protein maintenance and could even relate them to degradation phases. However, the small enrichment scores for each cluster and high false

discovery rate for each single gene function indicate that the gene functions I found do not relate directly to rhythmic PTR.

In summary, it is possible to draw some general conclusions regarding the nature of rhythmic PTR. The investigation of predicted degradation phases and phases of rhythmic RBP transcripts indicate that rhythmic PTR is more pronounced at specific times of the day. These times do not relate to the overall RNA phases, see Figure 2.2B. Furthermore, comparison between both organs reveals that rhythmic PTR is highly organ-specific. However, revealing further biologically motivated evidence for rhythmic PTR was not feasible. Functional RBP binding site were drowned in noise and gene function enrichment failed due to the small and specific gene set and possibly also to an incomplete gene function annotation. It seems only direct experimental evidence can support the modeling findings, as for example the investigation of single RBPs with rhythmic abundance or activity together with the RBPs' targets.



liver		kidney	
cell-cell adherens	11.38	protein transport	1.22
protein folding, chaperons	9.44	membrane fusion, SNARE	0.75
protein transport	8.45	calcium ion binding	0.71
ribosomal protein	5.64		
mitochondrion	4.33		
splicosome, mRNA splicing	3.88		
ATP-binding, nucleotide binding (18 - 6 h)	2.63	protein transport (6 - 18 h)	1.63
chaperone, protein folding (16 - 24 h)	1.89	vesicle mediated transport (12 - 0 h)	1.3
lipo protein (6 - 18 h)	1.41	mitochondrion (6-18 h)	0.99

Figure 2.12: **Rhythmic PTR is organ specific.** **A** Shown are histograms phases of RBP transcripts rhythmic in liver or kidney and phases of predicted degradation rates. Color code as in previous figure. **B** Illustration of strategy, how I tested for gene function enrichment using the DAVID Functional Annotation Tool. Table shows result for liver and kidney testing different gene sets. Phases are derived by dividing the gene set according to diverse ranges of predicted degradation phases, see Appendix C.5. Shown is most plausible phase range for the specific function. The low "cluster enrichment values" were accompanied by high enrichment q-values of each gene function (not shown here).

2.7.4 Shape of Time Series

Another approach implemented to investigate hints for rhythmic PTR is to look at the shape of the time series. A numeric simulation of the model from Section 2.5 reveals that rhythmic degradation changes the form of the RNA time series in a specific way. While rhythmic production described by cosine-function results again in a cosine-function in the RNA abundance, rhythmic degradation rate introduces narrow peaks and broad valleys to the time series, see Figure 2.13A. If interpreted as time-dependent functions these abundances can be approximated by Fourier series with a limited number of terms. A Fourier series of a pure cosine function consists only of the first term. More complicated functions as, for example, generated by a rhythmic degradation require more Fourier terms to be approximated, see Figure 2.13 A&B. In this way, RNA abundance mainly influenced by rhythmic production would be approximated by a Fourier series with the same number of Fourier terms as its rhythmic production, while rhythmic degradation would increase the number of Fourier terms.

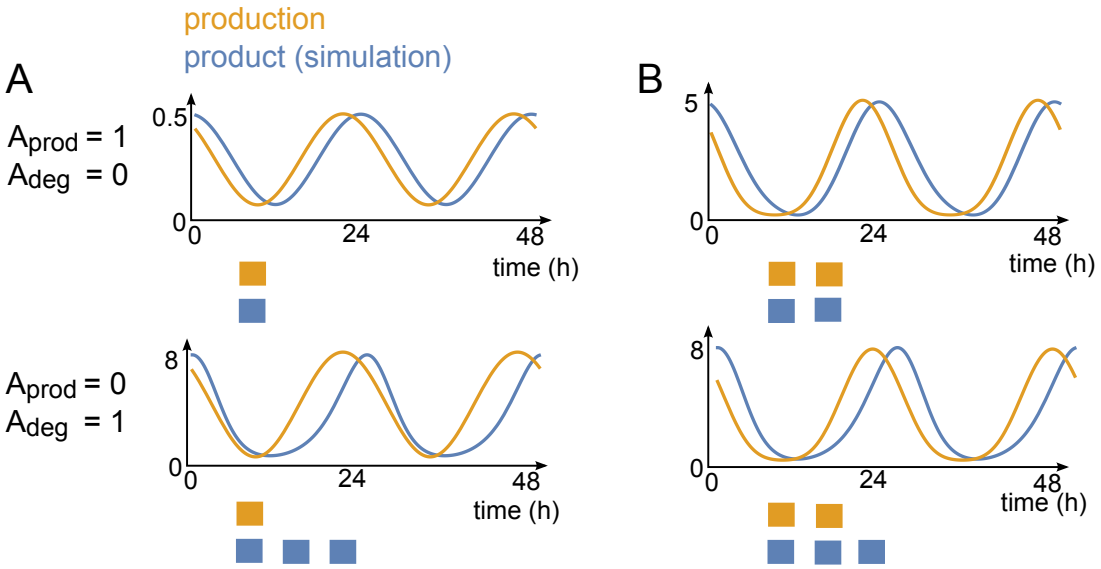


Figure 2.13: **Rhythmic degradation alters the shape of mRNA time series.** Shown are simulations of the model 2.9 in blue, in yellow for comparison the production rate, squares illustrate how many Fourier terms are needed to describe each function. **A** Production described by cosine function, **B** production described by a cosine production + 0.3 times its second harmonic.

Hence, I argue that if fitting of an RNA time series would need more Fourier terms than a fit of its transcriptional abundance this gene is likely to be tested positively to be under rhythmic PTR (methods see Appendix C.6). Accordingly, I first tested how many Fourier terms would be needed to describe transcriptional activity and secondly, if transcript abundance would be better fitted with a Fourier Series with one more term than its transcriptional activity. To compare different model fits I used the ANOVA-package implemented in R. The p-values of both tests, test of the time series shape and for rhythmic PTR, do not correlate. Also large relative amplitudes of the predicted degradation rate, an indicator for a strong influence of rhythmic PTR, only correlate with the test result testing for rhythmic PTR but not with the test for a different shape of the time series, see Figure 2.14A. However, both tests test different things. The p-value distribution of any test under the null-hypothesis is uniformly distributed. Hence, large p-values of independent tests are always uncorrelated. I therefore tested if hits of one test are enriched in the hits of the other test, see Appendix C.6. It turned out that this is true, with high significance, see Figure 2.14B.

Consequently, investigation of the shape of time series of transcripts returns the genes found to be under rhythmic degradation. However, not every gene with rhythmic PTR shows this characteristics in the time series. A possible reason for this may be the limited number of measured time points, which does not allow to fit Fourier series with many terms. Additionally, other kinetic influences, not captured by the simple model, might shape time series.

It is worth mentioning that the shape of time series altered by rhythmic degradation also influences the RNA's magnitude. This leads to higher magnitudes for RNA influenced by rhythmic degradation compared to magnitudes of RNA influenced by rhythmic production. We termed this phenomenon "magnitude effect" in the paper [179].

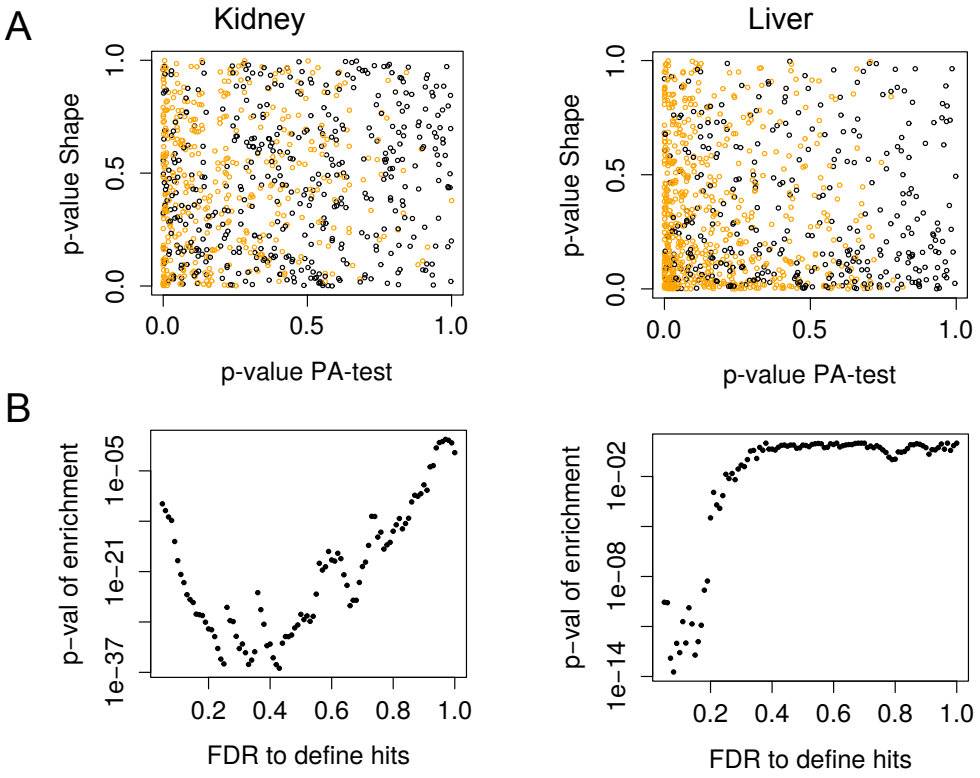


Figure 2.14: **Hits in genes with a different Fourier series fit for transcriptional activity and RNA abundance are enriched in genes with rhythmic PTR. A** Scatter plots of p-values of both tests. Orange indicate genes where relative amplitude of transcriptional activity A_{TRXact} is much smaller than relative amplitude of predicted degradation rate A_{deg}, *i.e.* $\frac{A_{\mathrm{deg}}}{A_{\mathrm{TRXact}}} \geq 1.5$ **B** p-value of enrichment test from genes with a different Fourier series fit for transcriptional activity and RNA abundance in genes with rhythmic PTR dependent on threshold to define hits in each test set.

2.8 Discussion of the First Part

In the preceding chapter, I answered the main question of what effect post-transcriptional regulation, constant or rhythmic, have on rhythms in mRNA abundance. First, in Section 2.1, I investigated how rhythms from transcription are transmitted to mRNA abundance. The analysis showed that every mRNA processing step dampens the rhythm. The longer it takes for processing, the greater the loss. Since mRNA half-lives introduce the longest time scales in mRNA processing, half-lives contribute most to rhythm loss. The presented mathematical relation between processing time and loss of rhythm allows the quantification of such effects. If a transcript is influenced by only one rhythm in its production, it can peak at most 6 hours (one quarter of a period) after the peak in production. The exact phase difference depends again on its processing time.

If mRNA abundance and transcriptional activity are compared on a genome-wide scale, one can observe several transcripts where mRNA peaks much later than 6 hours after its transcriptional peak activity. Furthermore, rhythm gain from transcriptional activity to mRNA abundance can be observed, that is, mRNA abundance shows a higher relative amplitude then mRNA transcriptional activity. These observations imply additional rhythms influencing the mRNA abundance. My model was kept general, allowing to systematically test diverse couples of potentially rhythmic processing steps. An obvious example for this is rhythmic degradation. I introduced a model describing rhythmic production and rhythmic degradation in Section 2.5 and presented an analytical solution together with an intuitive interpretation of the model solution. In Section 2.6 the model is re-interpreted and extended to test other candidates of rhythmic processing steps. The discussion of the model solution revealed that any phase relationship between mRNA abundance and transcriptional activity is now possible. For extreme cases, e.g., a phase shift of 12 hours between production and abundance, the second rhythmic process must, firstly, be the dominating oscillation, that is, it has to oscillate with a higher relative amplitude and, secondly, it must be uncoupled from transcription. After a careful discussion of the known post-transcriptional processes, I conclude that, with the information we have to date, only rhythmic degradation, cellular export or alternative splicing can explain the wide range of phase relationships between transcriptional activity and transcript abundance. In our publication we analyzed two specific data sets[179]. Transcripts that showed evidence of a second rhythmic processing step were enriched in targets of the RBP CIRBP and in mRNA with rhythmic poly(A) tail lengths. Poly(A) tail length is connected to its transcript's half-life and CIRBP is known to destabilize its target mRNA. Hence, both experimental findings point to rhythmic mRNA degradation.

If a transcript is influenced by rhythmic production and constant degradation there are strict limits for phase and amplitude relationships between mRNA abundance and transcriptional activity. This understanding allowed the development of a variety of tests whether a rhythm in mRNA abundance can be explained by the rhythm of its transcriptional activity alone. These tests are presented in Section 2.3 and 2.4. The most powerful of these is the so-called PA-test (production-abundance-test). It takes time series specific noise into account and provides a probability (p-value) to reject the null hypothesis of no additional rhythmic process. The test has, however, quite stringent data requirements: For each tested gene the half-life must

be independently measured at least twice. This is not provided for many genes. Therefore, I presented and discussed alternative ideas to test for rhythmic post-transcriptional regulation.

I used the modeling ideas together with the PA-test to investigate two data sets on mouse liver and kidney. This led to an estimate that 34% of circadian genes in liver are under rhythmic post-transcriptional control, while in kidney this portion is found to be much smaller, with only 18%. One reason for this discrepancy might stem from the longer regulatory sequences in liver RNA of circadian genes compared to circadian genes in kidney, a difference which was not observed between all expressed genes in liver and kidney, see Section 2.7.1. Longer regulatory sequences give rise to the possibility of harboring more RBP-binding sites and, hence, might provide more regulation possibilities. Furthermore, the predicted phases of the post-transcriptional processes and measured phases of rhythmic transcripts of RNA-binding proteins, showed no systematic correlation between the two organs. These observations may indicate that rhythmic post-transcriptional regulation is highly organ-specific. Notably, unusually long 3'UTR have been observed in the mammalian brain [225]. An investigation of circadian transcriptomes of the brain would be an intriguing application of the tool set presented herein.

I took another approach to verify genes with post-transcriptional control and investigated the shape of time-series, an idea based on numerical simulations of the model. Here, I could show an enrichment of genes where the shape of the time series indicate a rhythmic PTR in hits of genes positively tested for rhythmic PTR with the PA-test.

Unfortunately, the search for further independent evidence for rhythmic post-transcriptional regulation of the identified candidates in liver and kidney was not very fruitful. Neither RNA-binding site nor gene function enrichment analyses revealed further confirmation. Further experimental evidence might be necessary to verify the post-transcriptional control for the identified candidates. However, motif enrichment analysis is an art in itself, and further improvements may change the picture. One possibility for example would be to account for RNA accessibility of RBP-binding sites, that is, to exclude motifs that are masked by RNA secondary structure [226] into the previous analysis to reduce false-positives. Probably, other refinements of the proposed investigation methods are also possible.

3 Other Aspects of Rhythmic Post-Transcriptional Regulation

3.1 Can Harmonics be Generated by Post-Transcriptional Regulation?

In all of the previous results I left the period untouched and implicitly assumed that every transcript under rhythmic post-transcriptional regulation has a circadian period of about 24 hours. However, besides the circadian period, 12 hour rhythms were also observed in transcript abundances in mouse liver [106, 129, 227], although in only 1% of all circadian transcripts [106]. Cretenet *et al.* [228] investigated a pathway which activates with a 12-hour rhythm, the IRE1α pathway, in more depth. They not only showed 12-hour rhythms in mRNA abundances, but also that proteins oscillate in this shorter rhythm. Furthermore, they demonstrated that a disruption of the clock also diminishes the 12 hour-rhythms. They hypothesize the rhythmic pathway activation relates to a stress response of the endoplasmatic reticulum which is also reported to have a 12 hour rhythm. The findings of Cretenet *et al.*[228] not only demonstrate that 12 hour rhythms fulfill a biological function, but also that the circadian clock is needed for their generation.

A 12 hour rhythm is the second harmonic of a 24 hour rhythm, that means the frequency is twice as high as the original frequency. Which biological processes can generate this higher harmonic? At present an experimental proof is missing. Westermark *et al.*[229] showed that 12-hour rhythms can be generated by certain transcriptional activation patterns. If two transcription factors with a circadian abundance activate independently from each other the transcription of one gene, they can generate a second harmonic in the gene's transcript. The second harmonic is most pronounced if the abundance of the two transcription factors oscillate in anti-phase. Here, the coordination of the two rhythms with the same frequency gave rise to higher harmonics. Can rhythmic post-transcriptional regulation combined with rhythmic transcriptional activity, *i.e.* two rhythms influencing the transcript abundance, generate a 12 hour rhythm? In the following I want to use the previous model results to investigate this question.

Any rhythmic time series can be approximated with a Fourier Series, a linear combination of sines and cosines with whole-number multiples of the frequency:

$$x(t) = \frac{A_0}{2} + A_1 \cos(\omega t) + B_1 \sin(\omega t) + A_2 \cos(2\omega t) + B_2 \sin(2\omega t) + ... \tag{3.1}$$

Here, $\omega = \frac{2\pi}{24}\mathrm{h}^{-1}$ describes the angular frequency related to the circadian period, and A_0, A_1, B_1 ... are the Fourier coefficients, which determine the contribution of each sine or cosine function. We observe a second harmonic in the time series if the contribution of the first

frequency is smaller or equal compared to the contribution of the second, *i.e.* $A_1 \leq A_2$ and $B_1 \leq B_2$.

For a transcript which is rhythmically produced and rhythmically degraded I show in Appendix A.1.3 that the Fourier coefficients are proportional to 1 over the Fourier terms order, that is A_n $1/n$ and B_n $1/n$. Hence, the first Fourier term is larger than the second. Consequently, rhythmic production together with rhythmic degradation cannot generate a 12 hour rhythm. This is supported by extensive numerical investigation.

Rhythmic splicing, however, is able to generate a 12 hour rhythm as I will show in the following. An mRNA which is rhythmically produced and rhythmically spliced reads:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \mathrm{prod}_x(t) - \mathrm{splic}(t)x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \mathrm{splic}(t)x - \mathrm{deg}y, \\ \mathrm{prod}_x(t) &= k\left(1 + A_{\mathrm{prod}} \cos\left(\omega t - \phi_{\mathrm{prod}}\right)\right), \\ \mathrm{splic}_x(t) &= L\left(1 + A_{\mathrm{splic}} \cos\left(\omega t - \phi_{\mathrm{deg}}\right)\right), \\ \mathrm{deg}_y &= \gamma. \end{aligned} \tag{3.2}$$

Here, x describes the unspliced pre-RNA, y the mature mRNA. The time-dependent production and splicing rate is again described with a specific relative amplitude and a phase. The production term of y, the mature mRNA, consists of a product of two periodic functions: the rhythmic splicing rate and the rhythmic level of the unspliced pre-RNA x. This product gives rise to the 12 hour rhythm. In principle, a product of two pure cosine functions always oscillates with the doubled frequency, independent of their phase relation ship. This can be shown with the calculation of the product of two cosines with a phase difference ϕ

$$\cos(\omega t)\cos(\omega t - \phi) = \frac{1}{2}\left(\cos(2\omega t - \phi) + \cos(\phi)\right). \tag{3.3}$$

While the input oscillations oscillate with ω their product oscillates with 2ω, independent of their phase relationship. However, the reduction of the pre-RNA to a cosine function is misleading. Numerical investigation reveal that the generation of 12 hour rhythms is not as immanent as the calculation suggests, but, in contrast, strongly depends on the properties of production and splicing rates. Here, not only the relative amplitude of each rate and the phase difference between both rates play an important role, but also the shape of the rates influences the generation of 12 hour rhythms. If we assume that production and splicing rate follow cosine functions, then in order to observe 12 hour rhythms either the oscillation in splicing rate must be very strong or splicing peaks more than 12 hours later than the production, see Figure 3.1A. This result changes if we assume that the production is described by a rather pointy function, see Figure 3.1B and Appendix D.1. Such a pointy function has been observed both for kidney and for liver as exemplified in Section 2.2. In Figure 3.1B the parameter range which generates 12 hour rhythms for a pointy production rate is shown. Comparison with the previous result reveals that the graph is rotated. This means, in order to produce a 12 hour rhythm and assuming a pointy production rate, the phase difference between both rates must be smaller compared to a

system with a cosine-shaped production rate. However, a large relative amplitude in the splicing rate is still required.

Are these results biologically relevant? Are large phase differences between production and splicing plausible? What about a strong oscillation in the splicing rate? A large phase difference means that unspliced transcripts remain in the nucleus for a quarter day or longer. Most of the splicing, however, is coupled to the transcription [198]. Furthermore, it seems that the splicing machinery protects the pre-RNA from early nuclear degradation [230]. A long residence of unprocessed pre-RNA in the nucleus seems therefore to be unlikely. Hence, I expect the phase difference between production and splicing to be small, rather in the range of minutes than hours. Therefore, a large proportion of phase differences which would, in principle, generate 12 hour rhythms are not realistic. High relative amplitudes in the rates, on the other hand, are in principal possible. However, most circadian transcripts, an indicator for enzyme abundance and therefore also for rate amplitudes, have low relative amplitudes. Additionally, oscillations observed in mRNA can be easily lost in protein abundances due to the kinetics explained in Section 2.1. An increase of amplitude is possible due to several mechanisms, *e.g.* rhythmic protein degradation as discussed in this work to great extent. But this requires a finely tuned regulation of protein abundance.

Hence, the model results suggest that splicing can theoretically produce 12 hour rhythms. If we observe 12 hour rhythms due to splicing this process seems to be highly regulated as it poses high demands on the shape of the production rate and the splicing rate's relative amplitude. I argue that this would need a high evolutionary pressure to evolve, there would have to be a strong advantage of 12 hour rhythms over 24 hour rhythms.

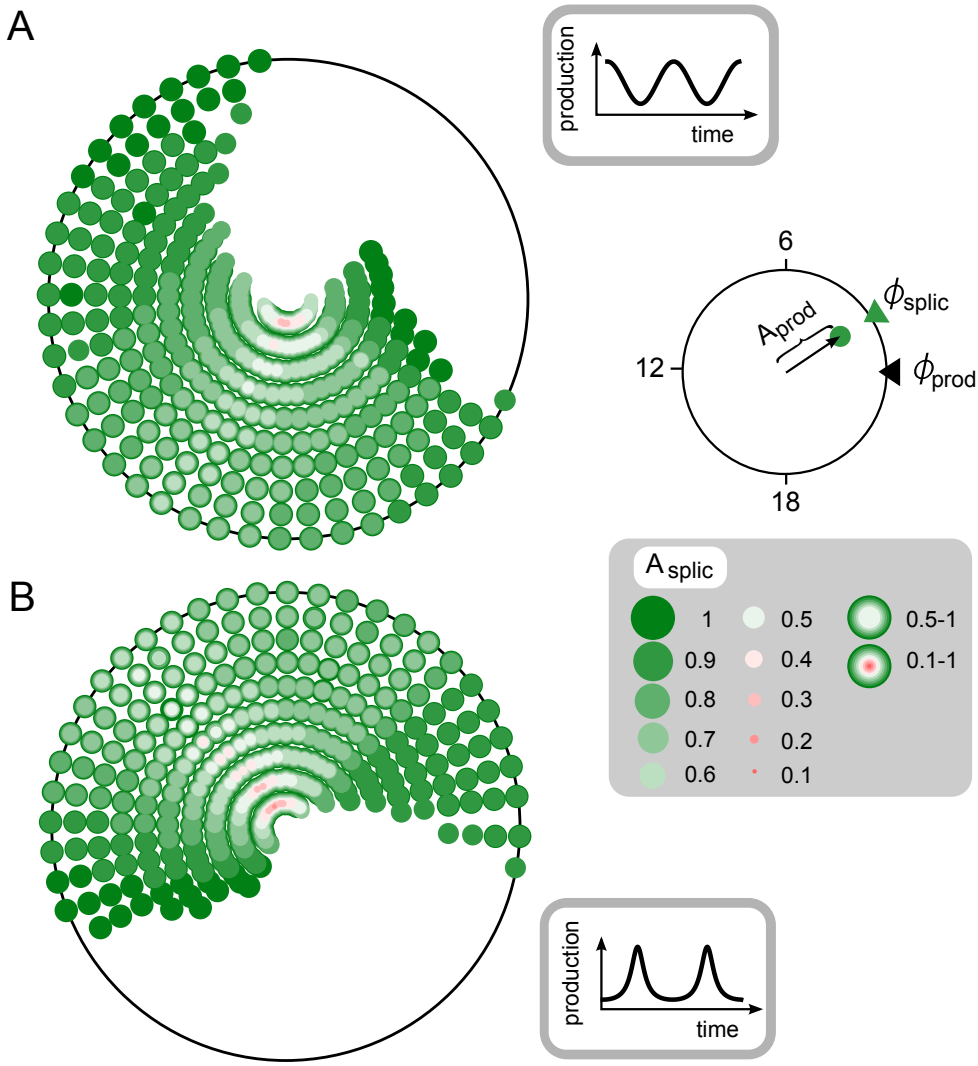


Figure 3.1: **Rhythmic splicing can produce 12 hour rhythms.** Shown are two systems for which the generation of 12 hour rhythms by rhythmic splicing were investigated. The large circles show the result, on the left is the legend. If a 12 hour rhythm was observed for certain parameters, a colored dot is plotted. All results are aligned to one production phase ϕ_{prod}. The position of the dot is determined by the relative amplitude of production A_{prod} and the splicing phase ϕ_{splic}. The splicing amplitude is encoded by color and size of the dot. They can stack onto each other and still be visible, see legend in gray box. **A** The model was simulated with a cosine shaped production. **B** The model was simulated with a pointy production, see Appendix D.1.

3.2 Post-Transcriptional Regulation of Clock Genes

The circadian core clock is a network of transcription-translational feedback loops. With its tuned inhibition and activation of gene expression the clock components oscillate with a period of about 24 hours. They are the molecular basis for any circadian regulation, including rhythmic PTR. However, core clock genes themselves are targets for rhythmic degradation, both at the transcript [159, 160, 161] and the protein level [231, 232]. How does a rhythmic degradation of a core clock component effect its oscillation properties, amplitude and phase? Is rhythmic PTR of core clock components able to change the period, a key feature of the circadian clock?

Rhythmic degradation within the core clock is not captured by the model presented in Section 2.5. Up to now I investigated only genes which do not feed back into a gene network system, see Figure 3.2A. A basic description of a network system which can produce oscillations is provided by the Goodwin model [57]. It provides a phenomenological description of a protein which suppresses the transcription of its own gene. It features the three important components that generate oscillations: negative feedback, delay and nonlinearity [56, 233]. The circadian clock is modeled by the Goodwin oscillator itself [234, 235] or by closely related models [236, 237, 238]. Consequently, the Goodwin model provides a good starting point to characterize the influence of rhythmic PTR on core clock components.

The Goodwin model is a 3-variable model with two activations and one inhibition. The synthesis and degradation rates are linear except of the inhibition rate, which takes the form of a sigmoidal curve, a Hill-curve. To simulate rhythmic PTR in the core clock I let the degradation rate of either one of the variables be rhythmic. This is achieved by introducing a cosine function with phase and amplitude which does not affect the average degradation rate, see Figure 3.2B, Appendix E.1. For each degradation phase and degradation amplitude I simulate the system and determine the oscillation properties, period, magnitude, relative amplitude and phase, of each system variable, see Appendix E.1 for details. In Figure 3.2C an example for a Goodwin oscillator with rhythmic degradation rate is plotted together with the original Goodwin oscillator with constant degradation rates.

Since it is well established that longer half-lives prolong the period of the Goodwin oscillator [234], I expected the period to be affected also by periodic half-lives. This could not be confirmed, instead the period remained stable for almost all parameters, see Figure 3.3. One exception was observed: If the degradation rate of the repressing species oscillated with an amplitude larger than 0.7 I observed a period doubling. All species then oscillated with twice the period, however this oscillation still contained the original period, see Figure 3.2D.

Apart from this exception, introducing a rhythmic degradation to the Goodwin model had a very stable effect, irrespective which variable was affected by rhythmic degradation. An increasing relative amplitude in the degradation rate resulted in an stronger overall oscillation of the whole system regardless of the degradation phase. This amplitude increase was very large, with up to a 2 fold increase for strongly oscillating degradation rates. This degradation phase independent amplitude increase contrasts with the results of clock-controlled genes, where only a specific phase range produced an amplitude boost, see Section 2.5. Increasing the degradation amplitude also increased the magnitude of all variables, see Figure 3.3. The degradation phase affected only the phase of the oscillating system. The relationship between degradation phase

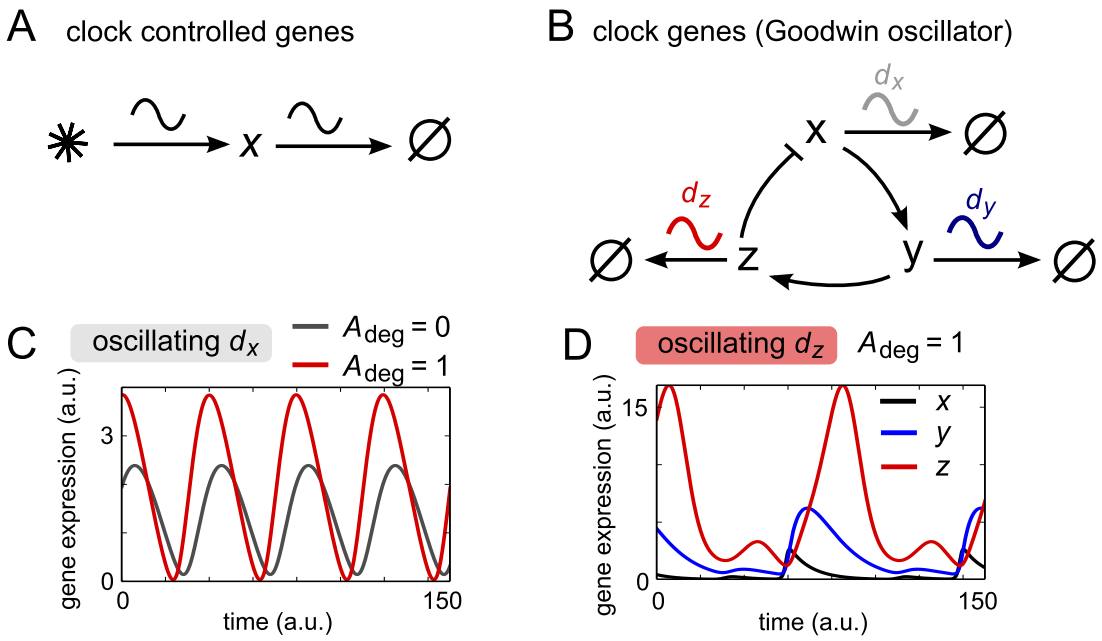


Figure 3.2: **Goodwin model with oscillating degradation rates as example for rhythmically degraded core clock genes.** For modeling details, equations and parameters, see Appendix E.1. **A** Modeling system of previous sections captures clock-controlled genes with rhythmic degradation. **B** Schematic description of the Goodwin oscillator with rhythmic degradation rates. **C** Simulation of Goodwin oscillator with and without rhythmic degradation. **D** High relative amplitudes in the degradation rate introduce periods twice as the original period. For modeling details, equations and parameters, see Appendix E.1.

and phase of the system is linear and the phase range of the system spans over all possible phases.

These results suggest that rhythmic degradation of core clock genes is beneficial to the core clock as it increases the overall amplitude and, hence, the biological impact of oscillations. Furthermore, rhythmic degradation could serve to easliy shift the phase of all core clock genes since there is a linear relationship between degradation phase and system phase. It seems, that oscillating degradation rates in this model system contribute to the overall energy of the oscillating system. However, the Goodwin oscillator is a toy-model which only consists of one negative feedback loop. The circadian clock network, on the other hand, contains several intertwined negative and positive feedback loops. Conclusions drawn from the Goodwin oscillator's properties are not necessary true for the circadian clock itself. For example, the fact that longer half-lives cause generally longer periods is true for the Goodwin model [234] but could not be confirmed for a more complex model of the circadian network [239]. Experimental evidence underlines this. A longer period of the core clock was observed either with a longer Cry1-mRNA half-life [161] or with a shortened Cry1-protein half-life [232].

Interestingly, downregulation or upregulation of degradation rate influencing factors - common experimental tools to investigate biological systems - primarily changes the average degradation rate, but not necessarily its rhythmic properties, relative amplitude and phase. To investigate only the rhythmicity of degradation one would need a tightly controlled down or upregulation at specific times of the day in a system which forgets quickly relative to circadian time scales. To my knowledge, such an experimental system has not been established. At present, it is not possible to disturb and, with that, investigate experimentally the rhythmicity of degradation rates and their contribution to the clock properties. Consequently, this problem can only be tackled by mathematical modeling.

In summary, for a rather simple model I find that when introducing oscillating degradation the system's period remains stable for most of the cases. Any rhythmicity in the degradation rate contributed to a stronger overall oscillation in all system's variables, which might be desirable for biological function. For more complex systems, these very general findings must be reviewed. In order to do that, one could introduce rhythmic degradation to an already established mathematical model of the circadian clock, for example in the model of Relogio *et al.* [239] or Woller *et al.* [240].

However, the proposed model, a Goodwin oscillator with modulated rates, should be discussed a little further. In this model I actually modulated the degradation rate externally. With this, the model describes an oscillating system driven by an oscillating degradation rate with the system's period. This means we are looking at an oscillator driven by an external force in resonance and this easily explains the contribution of oscillating degradation rates to the overall system's oscillation strength. Another implementation of this system is realized when the core-clock itself modulates the degradation rate. Hence, the degradation rates are modulated by one of the variables, for example $d_x = d_x(y)$.

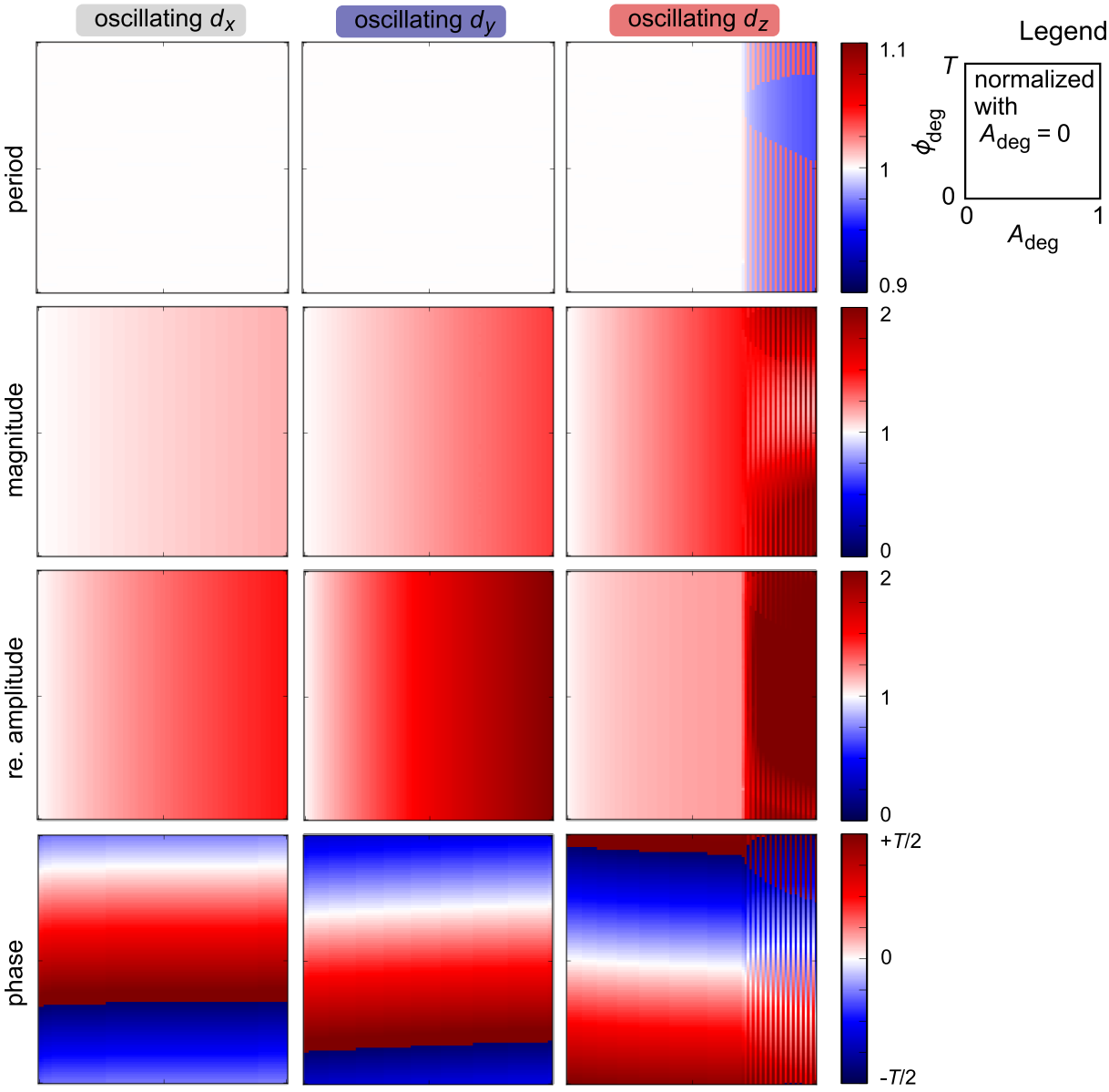


Figure 3.3: **Systematic analysis of rhythmic degradation in all degradation rates for Goodwin model.** For modeling details, equations and parameters, see Appendix E.1. Shown are normalized values. Period, magnitude and amplitude are normalized as follows: quantity(A_{deg})/quantity($A_{\text{deg}} = 0$). For the phase the phase difference is shown: $\phi(A - \text{deg} = 0) - \phi(A - \text{deg})$. When the system oscillates with a period $2T$, *i.e.* $A_{\text{deg},z} > 0.7$ the parameter estimation does not work, since it relies on the short period, see Appendix E.1. Color code: White represents the parameters of the system without oscillation. T =period of the system without oscillating degradation.

3.3 Aging of Molecules - Poly(A) Tail and Oxidized Proteins

I started this journey through rhythms and post-transcriptional regulation with the fundamental statement that time matters for rhythms and long-lived molecules will loose their rhythms. In this last section I want to partially withdraw this statement and show that molecules do not completely loose their rhythms, but rather remember their history. If the age of a molecule is measurable it is possible to see traces of rhythmic production also in long-lived molecules.

I describe the aging of a molecule by a Partial Differential Equation (PDE), a direct extension of the ODE model describing rhythmic production and rhythmic degradation from Section 2.5. In this chapter, I will introduce this model and characterize its dynamics. I will present two examples where the cryptic concept of a molecule's age can be related to actual data. And lastly, I will demonstrate how a molecule remembers its history and how the PDE model can be used to predict other, yet unmeasurable, dynamic quantities.

3.3.1 The Aging of Molecules - a PDE with a Physical Background

When we look at a person we can easily make a good guess how old he or she is by taking into account several signs of aging such as gray hair or wrinkles. This concept of age does not translate directly to biological molecules since neither proteins nor mRNAs get wrinkles, but in both cases we can find different aspects which can be an indicator for a molecule's age. In the case of proteins it is known that long-lived proteins accumulate damage, mainly caused by reactive oxygen species (ROS). mRNAs however, receive after their transcription a stabilizing element, the poly(A) tail, a chain of about 250 adenosides. This chain is eroded during an mRNA life time. Roughly speaking an "aged" protein is then a protein which accumulated damage and an aged mRNA is an mRNA where the poly(A) tail is shortened. However, both processes, accumulation of damage in proteins and erodation of the poly(A) tail in mRNA, do not always proceed linearly in time as the term "aging" might suggest. For example there might be times during the day with an increased ROS production and consequently a higher protein damage. Similarly, it has been suggested that erodation or deadenylation of a poly(A) tail happens quite quickly for the first adenosides but slows down for the last ones it. In both cases the "aging" of either proteins or mRNA occurs faster or slower depending on the condition and background.

The McKendrick [241, 242] or von-Foerster [243] equation describes the aging of a population. I use this model to describe "aging" molecules and modify it in such a way that the molecule's aging do not occur linearly in time. For simplicity, and to avoid the poor analogy of "aged" molecules, I will refer from now on to "damaged" molecules. This can be directly translated into damaged proteins. To translate this concept to mRNA with shortened poly(A) tail we calculate the mRNA with a "damaged" tail, a newly synthesized mRNA with a full poly(A) tail has no damage.

In the modified McKendrick model a molecule is characterized by two variables, time t and damage d. It can be represented in a three-dimensional graph, see Figure 3.4A. Let then $x(t, d)$ be the number of molecules of damage $(d, d + \mathrm{d}d)$ at time t. Hence, $x(t, d)$ is a density function in the damage d. An integration by the whole damage range gives the total concentration of

molecules X with all possible damages at a specific time t

$$X(t) = \int_{d=0}^{\infty} x(t,d)\ \mathrm{d}d. \tag{3.4}$$

The concentration of molecules within a certain damage range $D+\Delta d$ is given by an integration over a certain range of damage

$$X(t,D) = \int_{D}^{D+\Delta d} x(t,d)\ \mathrm{d}d. \tag{3.5}$$

The fact that the modeling variable actually refers only through integration to a physical property, molecule concentration, requires a careful derivation of the rates. This and a mathematical justification of the model is presented in Appendix B.1 and B.2. However, to translate from $x(t,d)$ to $X(t,D)$ in an intuitive way is rather straightforward, despite the not informative integration. Figure 3.4B illustrates the integration and hence the intuitive understanding for two cases. In short, $x(t,d)$ is linear to $X(t,D)$ for sufficiently small Δd. Using this linear relationship it is possible to find a translation from $x(t,d)$ to $X(t,D)$. To avoid cluttered text I will use in the following the term molecule concentration for $x(t,d)$, although technically speaking it is a molecule concentration density.

A molecule which is produced, damaged during its life time and degraded is described with the PDE

$$\begin{aligned} \frac{\partial x(t,d)}{\partial t} &= -q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d} - v(t,d)x(t,d). \\ x(t,0) &= \frac{k(t)}{q(t,0)} \end{aligned} \tag{3.6}$$

Here, $k(t)$ is the production rate of newly synthesized proteins without damage, it characterizes the initial condition $x(t,d=0)$. The rates $v(t,d)$ and $q(t,d)$ refer to molecule degradation and damage accumulation, respectively. Since I want to describe a system under the influence of the circadian clock all rates contain a time dependent, periodic function which describes the oscillating variation of the rates. Degradation and damage accumulation can additionally depend on damage. This can reflect different scenarios: For example, in the case of protein oxidation, highly damaged proteins are more likely to be recognized for degradation than proteins with lower damage. In this case the degradation rate increases with damage. In the same manner the damage accumulation rate could depend on damage as I will show later. The actual mathematical formulation of this damage dependent term depends on the biological situation. In Appendix B.2 I discuss different scenarios for damage dependent degradation and damage accumulation rates and suggest some mathematical description.

The Equation 3.6, known from population dynamics, is also well known in physics in a very different context. In physics, this equation is used to describe the gas flux through a tube, its name is then “transport equation”. In the following I want to use this physical interpretation to discuss the different influences of the rates. This will, hopefully, provide us with a very intuitive understanding.

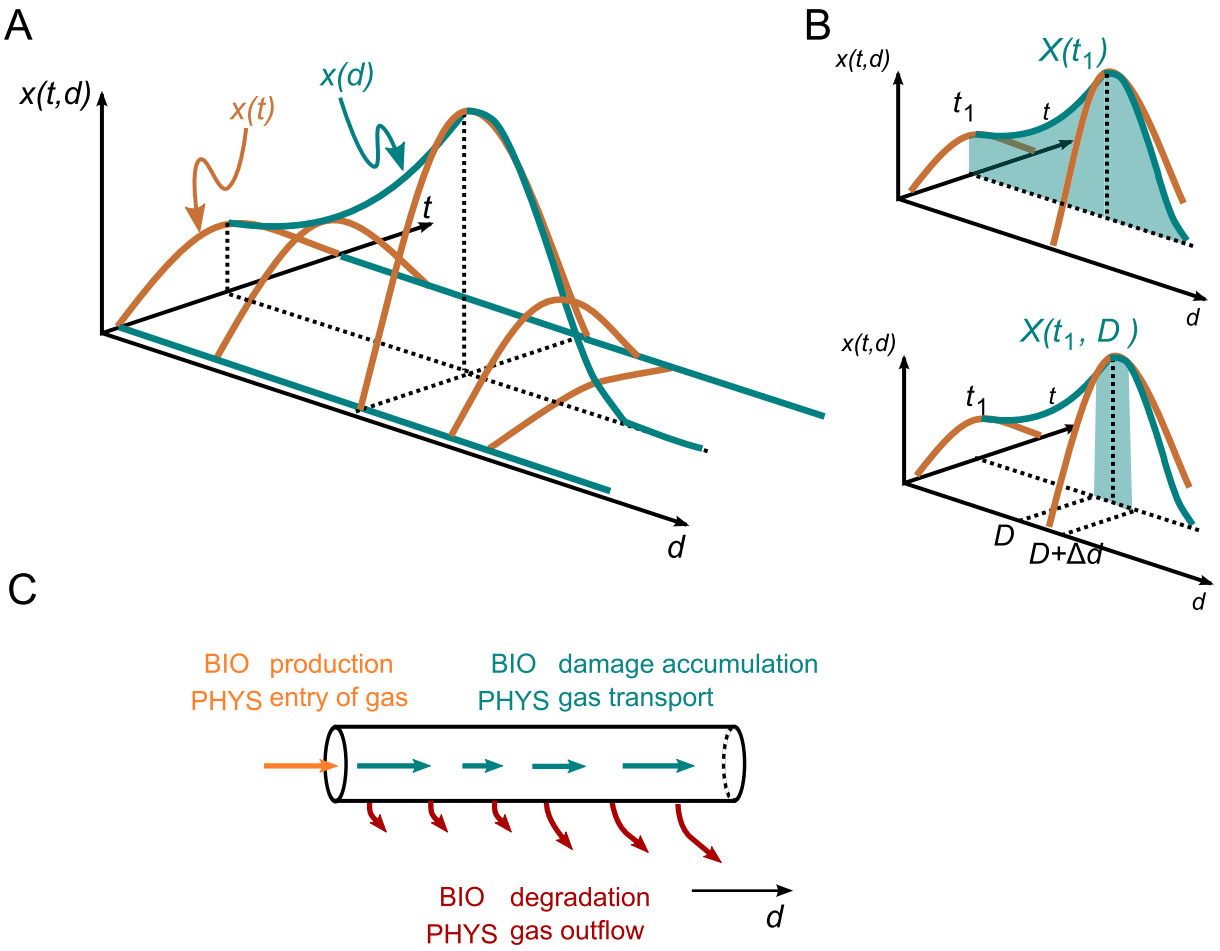


Figure 3.4: **A PDE describing damage accumulation of molecules. A** Schematic solution of $x(t, d)$. The function has two variables and is therefore depicted with three dimensions. Blue lines indicate damage dependence $x(d)$ at fixed time points, brown lines indicate time dependence $x(t)$ for fixed d. Use dashed lines for orientation. **B** Integration over d transforms molecule concentration density $x(t, d)$ to molecule concentration, either, when integrated over all d to $X(t)$ or, when integrated over a small range Δd to $X(D, t)$. Blue shaded area represents the integral $X(t) = \int_0^\infty x(t, d)\mathrm{d}d$ or $X(t, D) = \int_D^{D+\Delta d} x(t, d)\mathrm{d}d$ for a fixed time point t_1.**C** Interpretation of PDE as transport equation. The figure illustrates how different biological rates (BIO) translate to properties of gas transport in a tube (PHYS). Shown is a snapshot in time. The different arrow lengths illustrate different rate strengths at different d.

In this picture, the concentration of molecules $x(t, d)$ at time t and damage d is the gas density at time t and position d in the tube, see Figure 3.4C. This gas is poured into the tube only at one end of the tube. During this pouring the gas density is varied with a time-dependent rate. In this way, there are packages of highly dense gas and packages of low density gas produced at the opening and then further transported. In our damaged molecule system, the pouring of gas represents the production rate $k(t)$ of molecules with zero damage. The gas is further transported through the tube with small pumps, with a pump placed on each point along the tube. In this way different speeds of transport can be enforced. Furthermore, all pumps are controlled by one clock which dictates the same periodic rhythm to all pumps. If the small pumps have different power outputs they locally change the gas density. The time dependency of gas transport, however, can not generate a change in gas density since all pumps are controlled together. The time dependency of transport rate only changes the overall transport. Transport changes the position of gas, in the molecule image this is the damage of a molecule, hence the transport reflects the rate of damage accumulation. The last rate to be explained is the degradation rate. This rate can be imagined as little holes in the tube where gas escapes. Again, the hole size and therefore the escape rate can depend on the position on the tube. But all holes together are controlled by one clock which can close and open the holes. Also this rate is able to change the gas concentration, in this case both time and position dependent. However, degradation rate can only decrease gas density because it removes gas. In contrast to that the transport rate is able to both increase and decrease locally gas density when gas is pooled or thinned out.

It is worth mentioning that the model is only analytically solvable if damage accumulation and degradation rate do not depend on damage, see Appendix B.3. Hence, I will rely on numerical solutions. In Figure 3.5 some numerical simulations for different rates are shown.

This model describing a molecule concentration with two features, time and a property acquired over time is very general. Also the two very different examples which I will introduce in the following in more detail underline the model's generality. In the first example the model is used to describe long-lived proteins and their oxidation, the second describes the poly(A) tail of mRNA.

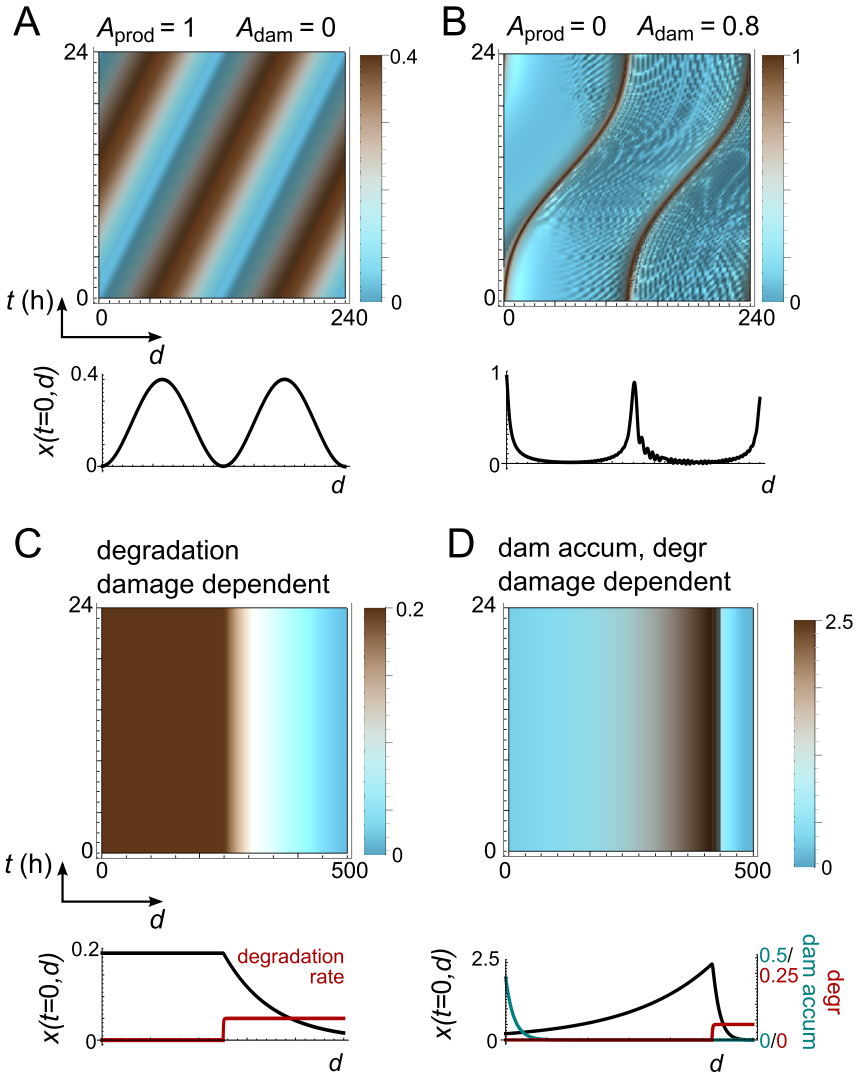


Figure 3.5: **Different numerical simulations of the PDE Equation 3.6**. For all graphs $x(t, d)$ as color-coded area together with a corresponding line profile (black) at time $t = 0$ is shown. **A&B** Time-dependent (rhythmic) properties of production $k(t)$ and damage accumulation rate $q(t)$ are varied. Degradation equals zero for the shown variable range, $v(t, d) = 0$. Damage accumulation is independent from damage, $q(t, d) = q(t)$. A_{prod} and A_{dam} define relative amplitudes of production and damage accumulation rate, respectively. In B numerical errors occur due to S-shaped sharp structures in the d-direction of the function. These errors visualize as interference pattern in the graph of $x(t, d)$. **C&D** All rates are time-independent. **C** Degradation rate is modeled as a step-function, see red line profile below. **D** Damage accumulation is proportional to an exponential function (blue in line profile), degradation as in C.

3.3.2 Oxidized Proteins – Rhythms are Conserved

Proteins, in contrast to mRNAs, have much longer half-life, for proteins a median of 46 hours compared to 9 hours for mRNA was detected [213]. Long half-lives destroy rhythms. Hence, despite their rhythmic production many proteins possibly do not oscillate in their abundance. Proteins however are exposed to reactive oxygen species (ROS), highly reactive molecules which are a by-product from mitochondrial respiration. These free radicals attack proteins in different ways. The most frequent alteration of attacked proteins is introducing a carbonyl group [244]. These alterations can be detected using an immunoblot technique (OxyBlot by Millipore). This makes it possible to measure the concentration of damaged proteins.

With that we now have all ingredients to implement the model: the two variables time t and damage d given by the concentration of proteins with carbonyl groups characterize the protein concentration $x(t, d)$. A simulation of long-lived proteins with a rhythmic production demonstrates that the protein concentration is almost constant if we look at the total proteins content whilst neglecting the damage variable. However, proteins with a certain damage, the measured carbonyl groups, still show the conserved rhythm in production rate, see Figure 3.3.2.

Rhythmic production can cause a rhythm in damaged proteins. Desvergne *et al.* [245] observed rhythms in proteins with carbonyl groups in two cell lines. They monitored all carbonylated proteins using OxyBlot. However, production of all proteins does not oscillate, especially not with one phase. This deems it very unlikely that rhythmic production caused the observed rhythms in damaged proteins. In addtion, an oscillating degradation rate, oscillating damage accumulation or a combination of both can generate oscillating levels of damaged proteins. Biological evidence is found for all these cases. Desvergne *et al.* [245] demonstrated that key components of the 20s proteasome, the proteasome which removes damaged proteins [246], oscillate with a circadian rhythm and hence the degradation rate oscillates. Mitochondrial activity and, with that, the ROS production and damage accumulation rate is controlled by the circadian clock [113, 114]. This could be caused by oscillating mitochondrial key enzymes [115, 114] as well as an oscillating mitochondrial dynamic [113] which is correlated with mitochondrial activity [116]. With further information on rate constants the proposed model Equation 3.6 can be used to discriminate these different influences.

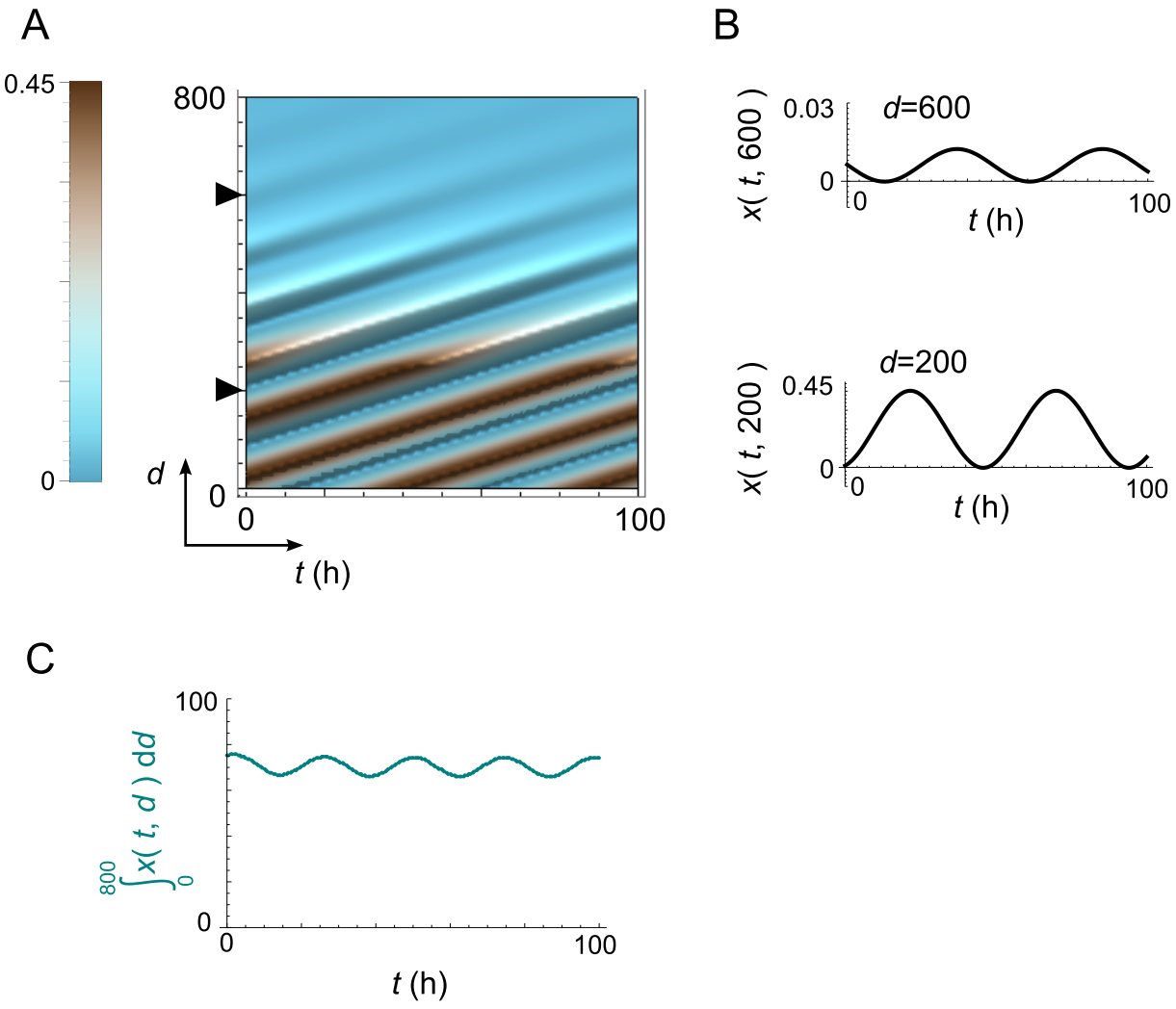


Figure 3.6: **Rhythms of production are conserved in long-lived proteins. A** Numerical solution of the PDE Equation 3.6. The PDE is modeled with an oscillating production ($A_{\text{prod}} = 1$), constant damage accumulation and a time-independent degradation rate with a step-function in d-direction, the same function as in Figure 3.5C & D. Black triangles indicate position of line profiles in B. **B** Two line profiles of the numeric solution at specific values of d. These line profiles illustrate a measurement of the level of oxidized proteins. **C** Integration of the numerical simulation in A. It illustrates the time-dependent level of of all proteins irrespective of their accumulated damage.

3.3.3 Deadenylation of Poly(A) Tail

Another example to which we can apply the model characterizes a post-transcriptional process, deadenylation of poly(A) tails. After transcription the poly(A) tail, in mammals a chain of usually 250 adenosides [247], is added to the pre-mRNA. This stabilizing element at the 3'-end of the mRNA is successively shortened in the cytosol by several deadenylases. The orchestration of the diverse deadenylases is not yet fully understood but it seems that different deadenylases successively attack the poly(A) tail [248, 187] and hence the poly(A) deadenylation rate depends on the poly(A) tail length. Once the poly(A) tail is removed from the mRNA the whole mRNA is attacked usually by exonucleases and quickly degraded. The half-life of an mRNA is therefore mainly determined by the deadenylation of the poly(A) tail. Furthermore, only mRNA with sufficient long poly(A) tails can be efficiently translated [249]. With that the length of the poly(A) tail dictates two key features of the mRNA, namely for how long and how efficiently can the mRNA fulfill its function *i.e.* being translated. However, the whole process of deadenylation is incompletely understood.

How can we elucidate the deadenylation process and its involved deadenylases in more detail? A big step in this direction is to look at deadenylation rates. Qualitatively changes in this rate over the length of the poly(A) tail or over the course of a day provides hints that the tail is attacked by different deadenylases or that deadenylases are controlled in a time-dependent manner. To my knowledge, an experimental approach to measure genome-wide deadenylation rates in vivo is not feasible at the moment. It is only possible to measure deadenylation rates of radio-labeled bulks of different RNA or deadenylation rates of single RNA using qPCR-based methods [250, 251]. However, we can describe the deadenylation of mRNA with the PDE model and use this with new sequencing-data to predict deadenylation rates in a genome-wide manner as I will outline in the following.

PDE-model Describing Poly(A) Tail Deadenylation

I will first translate this process of deadenylation into the PDE-model. In this framework, $x(t, d)$ describes the mRNA concentration with d already deadenylated adenosides, a newly synthesized mRNA has zero deadenylated adenosides. This means d counts as removed adenosides. Newly synthesized mRNA have a very narrow poly(A) tail length distribution of about 250 adenosides. This allows us to introduce a new variable d' in the model solution in order to translate to the actual length of the poly(A) tail, a more accessible variable. The new variable d' is determined by the variable transformation

$$d' = 250 - d. \tag{3.7}$$

In the model, the production rate $k(t)$ refers to the transcript synthesis or transcriptional activity. The damage accumulation rate $q(t)$ describes the deadenylation rate and the degradation rate the mRNA removal by exonucleases. Transcription of specific mRNA can exhibit a circadian rhythm, accordingly the production rate is time dependent. Degradation rate, the removal of the mRNA by exonucleases, depends strongly on the tail length. Only when the poly(A) tail length is removed, the mRNA is degraded. If there are exonucleases with a circadian rhythm is not known yet, however I deem it unlikely since this would mean accumulation of mRNA with no tail and hence nonfunctional mRNA with a circadian manner. To my knowledge, this has

not been observed. Consequently, the degradation rate is described well by a step function with a steep increase for $d \approx 250$ or $d' \approx 0$ and has no time-dependence.

Since deadenylation is performed by different proteins which successively attack the poly(A) tail, possibly each at a different rate, the deadenylation rate depends on the poly(A) tail length d'. The deadenylation rate can also exhibit a circadian rhythm due to circadian activity of Nocturnin, a deadenylase with cycling gene expression in various tissues [166]. Hence, the deadenylation rate depends on the poly(A) tail length d' and on time. It is not yet known if Nocturnin's activity requires a specific poly(A) tail length. If this was the case, the mathematical descriptions requires a function dependent on poly(A) tail length coupled to the time, best described by an amplitude which depends on the tail length d'. Kojima *et al.* [189] identified mRNA with long (60-250 adenosides) and short (<60 adenosides) poly(A) tails over the course of a day. Even with this coarse distinction they identified 237 mRNA with a rhythm in their short/long ratio of poly(A) tail length. Obviously, there is some influence of the circadian clock on poly(A) tail length control. As we learned from the previous example, an oscillating production rate results in an oscillation of mRNA with a specific poly(A) tail length. An oscillating deadenylation rate that additionally depend on length causes rhythmic profiles in poly(A) tail length. Although I discussed in detail time-dependent rates, for many mRNA no time-dependence in any of the rates is required. The model can serve to describe a time-independent, non-circadian deadenlytation aswell.

I will now use this PDE model to predict deadenylation rates.

Prediction of Deadenlytation Rates from Poly(A) Tail Sequencing

The basis of predicting genome-wide deadenylation rates is provided by an advancement in RNA sequencing. In 2014, two methods were published which sequenced the poly(A) tail. Up to this date, sequencing of long homopolymeric sequences (longer than 30 nucleotides) was not possible. The sequencing of long stretches of one nucleotide, adenosides for example, contain reading errors. The possibility for errors increases with the length. This makes it impossible to define the correct end of a long homopolymeric sequence. However, Chang *et al.* [252] and Subtelny *et al.* [253] independently found two methods called TAIL-seq and PAL-seq to overcome this problem. TAIL-seq uses a machine-learning based method to interpret the inaccurate signal [252], in PAL-seq the beginning and end of a poly(A) tail is chemically tagged [253]. With these methods both groups were able to measure distributions of poly(A) tail lengths in a genome-wide fashion.

In the model description the poly(A) tail length distribution is described by the solution $x(t, d)$ of the model. Inferring the deadenylation rate from the sequencing data, TAIL-seq or PAL-seq, poses a so-called inverse problem. What does this mean?

Usually, a model solution $x(t, d)$ is defined by its input parameters, here the production $k(t)$, deadenylation $q(t, d)$ and degradation rate $v(t, d)$. One can say, the model input parameters cause the model solution. To identify the deadenylation rate we have to inverse this problem. With the now known model solution, the information on poly(A) tail lengths, we want to recalculate one of the input parameters, the deadenylation rate. In order to obtain a unique solution we have to neglect the influence of other parameters. This can be done with two, biologically reasonable, assumptions. First, only mRNA with a short poly(A) tail, *e.g.* <20 nucleotides,

gets degraded, consequently, the degradation rate does not affect the mRNA for most of the poly(A) tail length. The second assumption demands that the poly(A) tail length distribution changes only slowly over time. Deadenylation is fast compared to circadian time scales. This neglects the time dependence of all rates. An investigation of oscillating deadenylation rates, hence time-dependent rates, would require a time series of snapshots of poly(A) tail length distributions. Each snapshot is analyzed independently and a comparison between deadenylation rates of different time points would reveal oscillations. Up to now, TAIL-seq of a time series is not available, but I know from personal communication that two groups are working on this problem.

If all these assumptions are fulfilled we can recalculate the deadenylation rate up to one linear factor, which accounts for the production rate. Mathematically this is done by a discretization of the solution and the model. Comparison of both results in a linear system of equations from which the deadenylation rate in dependence of the poly(A) tail length is calculated. In Appendix B.4 the method is outlined in more detail. Figure 3.7B shows a very simple simulation of some data. Interestingly, already the simple assumption on the deadenylation rate, namely

$$q(d)\ e^{-0.1d}, \tag{3.8}$$

can reproduce the appearance of already published data, see Fig 3.7A[1]. In Figure 3.7C the deadenylation rate is shown twice, once the one used for the simulation in Figure 3.7B and secondly, the one I calculated from the simulated data using the method described above. Both deadenylation rates agree very well proving the reliability of the proposed method.

[1]Reprinted from Molecular Cell, 53/6, Hyeshik Chang, Jaechul Lim, Minju Ha, V. Narry Kim, TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3' End Modifications, 6, 2014, with permission from Elsevier.

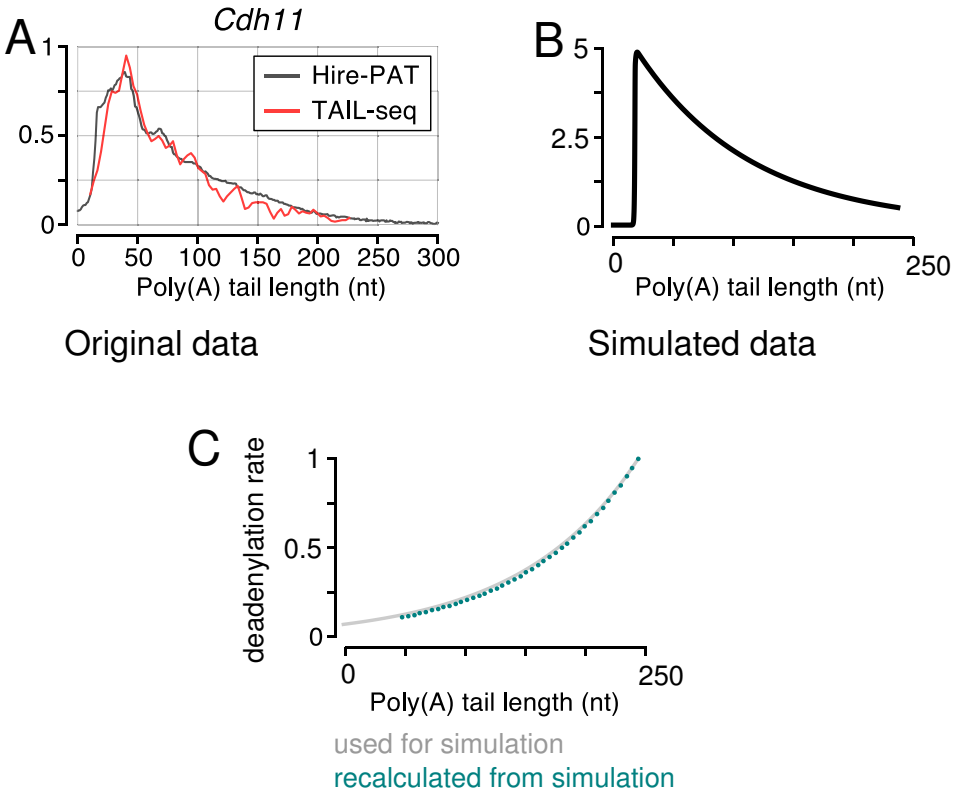


Figure 3.7: **Prediction of deadenylation rates.** **A** Measured poly(A) tail length distribution for Cdh11, reproduced from [252]. **B** Simulated data with time-independent production, damage accumulation and degradation rate. Damage accumulation $q(t) = 5e^{-0.1d}$, degradation is modeled as step function similar to the one in Figure 3.5C&D, see also Appendix B.2 for details. **C** Deadenylation rate $q(d)$. Gray: $q(d)$ used for simulation shown in panel B. Blue: $q(d)$ recalculated from simulation by using the inverse method, see Appendix B.4 for details.

3.4 Discussion of the Second Part

In the second part of this thesis I analyzed more specific aspects of post-transcriptional regulation in a hypothesis-driven approach. Can rhythmic splicing generate 12-hour periods? What happens if the core clock experiences rhythmic degradation? And finally, I extend my model to a partial differential equation model that accounts for an "aging" process of molecules.

It had been suggested that 12-hour rhythms can serve a biological function within a circadian context [228]. Using a two component model I found that rhythmic splicing is able to generate 12-hour rhythms in abundance of mature mRNA from circadian pre-mRNA production profiles. However, this poses stringent requirements on the ryhthms in transcription, *i.e.* transcription must oscillate with a very high relative amplitude. Such amplitudes are only observed in a hand-full of genes in liver and kidney; many of them are core clock members, see also Figure 2.3 for which, to my knowledge, no such harmonic expression patterns have been reported.

To investigate rhythmic degradation of core clock genes I introduced rhythmic degradation rates to a Goodwin oscillator. A systematic analysis showed that increasing rhythms in one of the degradation rates can strongly increase the amplitudes of all variables, driving the system in resonance. Thus, rhythmic degradation could well be involved in the observed exceptionally strong amplitudes of the core clock.

Interestingly, a rhythmic degradation of an inhibiting component of the Goodwin oscillator can yield periods twice as large as the original period, pointing towards a "period doubling bifurcation". Notably, such period doubling bifurcations had been suggested to have accompanied the evolution from primordial clocks to the extant circadian clocks during lengthening of earth's rotation period in early evolution [44]. The observed effect of rhythmic degradation may thus serve as an interesting starting point for further theoretical investigations. An interesting next extension of the model could be to let one of Goodwin variables directly modulate a degradation rate.

In the last section, I revisited the model of rhythmic production and rhythmic degradation and expanded it by another variable, the age or rather the damage of a molecule. This PDE describes the development of a molecule during its life. The PDE is described by three rates, molecule production, damage accumulation and finally molecule degradation. I discuss the influences of the rates with the help of an intuitive interpretation as a transport equation. With different mathematical formulation of the rates, also outlined in Appendix B.2, the PDE can be applied to describe different biological processes. In this thesis, I presented two distinct applications.

First, I described proteins which accumulate damage through accumulation of oxidation. Here, I showed that rhythms of a rhythmic production are lost if one observes the whole population, but are still pronounced when specifically looking at damaged, e.g. oxidized, proteins. This result implies that experimental observations of non-oscillatory protein levels should be taken with a grain of salt, since an oscillation in damaged fractions of the protein may still serve biological functions.

The second application of the PDE model is a description of poly(A) tail length development. This description provided the basis to calculate yet unmeasured genome-wide deadenylation rates using TAIL-seq data. In this thesis I provided a proof of concept and presented an example based on a numerical solution of the PDE model. This will be used to interprete upcoming TAIL-seq

data from the laboratory of Carla Green.

4 Concluding Remarks and Outlook

This thesis is based on one very simple modeling motif, rhythmic production and rhythmic degradation. One can think of it as a building block. This building block can be tuned to also describe constant production or degradation by switching rhythms on and off, simply by setting relative amplitudes to zero.

An analytical solution, an equation, together with its intuitive vector representation was presented. This is useful for a systematic analysis of parameters or fitting the model to data. The intuitive interpretation allows also scientists which are not trained in mathematics to work with this model. Furthermore, it facilitates a quick inspection of scientific ideas.

This building block was further varied to examine different aspects of rhythmic post-transcriptional processes. First, I varied the interpretation of the model. The same model described a diverse range of biological processes, such as rhythmic transcription, rhythmic splicing or rhythmic mRNA degradation. This could be achieved because the main motivation of the modeling approach was to investigate a very general idea, that is how rhythms in these very different processes affect mRNA abundance. This generality is in contrast to the idea of asking very specific questions regarding biological processes, such as how protein x influences the splicing efficiency.

I further varied the main building block and joined several of these blocks together. With this I obtained a modeling system describing more than one species, allowing the investigation of arbitrary phase relationships between transcriptional activity and mRNA abundance, and how rhythmic splicing can in theory, account for higher harmonics of mRNA abundance in the context of circadian transcription.

The PDE describing aging of molecules is also a variation of the main building block. Here, a second dimension besides time is introduced, the age of a molecule. This second dimension provieds the life of a molecule assigning different life stages to the molecule. Folding together this dimension and the life stages to the binary molecule's existence or non-existence reduces the PDE to our main building block. Mathematically, this is achieved by integration over all ages as described in Section 3.3.

Interestingly, also the modification of the Goodwin oscillator is inspired by the core modeling motif. As in the core motif I let the degradation rate be rhythmic by modeling them with cosine functions.

I applied this set of modeling tools to explain and analyze different aspects of rhythmic PTR, such as the extend of rhythmic PTR, which rhythmic post-transcriptional processes can contribute to the observed data and many more as explained in the discussion sections at the end of each chapter. I concentrated on post-transcriptional processes because the life of an mRNA is extensively covered by the diverse sequencing techniques. The birth of an mRNA can be monitored using ChIP-seq of polymerase II, GRO-seq and Nascent-seq. Transcript abundance is readily accessible by common RNA sequencing. Today, aspects of the mRNA decay is now

available by monitoring the poly(A)tail length through TAIL-seq or PAL-seq. All of these techniques work genome-wide and, with that, give a comprehensive insight into the lives of all expressed mRNAs. This data situation provides a rewarding ground to ask and answer specific questions using modeling.

However, it is equally possible to find other interpretations for the presented models. One obvious interpretation is to describe not rhythmic production and degradation of mRNA but of proteins or any other molecules. Also different binary states of molecules could be described by this model, for example an mRNA inside or outside of the nucleus. The rates can be reinterpreted. Instead of degradation, one can think of a permanent modification, for example RNA editing. It should also be tested how well the model can serve as a description of rhythmic reversible modifications, for example protein phosphorylation. In addition, the PDE model can probably be used to explain and analyze many more processes than the two presented in this thesis. The fact that the two presented processes describe very different biological processes underlines this.

Taken together: the model's modularity on one hand and the very general description of rhythmic processes on the other hand, makes the model a powerful tool to describe a diverse range of rhythmic biological processes, some of them outlined in this thesis, many more yet to be found. Hence, this thesis does not only provide a comprehensive analysis of rhythms in post-transcriptional processes, summarized in Section 2.8 and 3.4, but also provides some tools to handle models build out of the main building block and its variations in future. I expect that the analytical solution of a model with rhythmic production and rhythmic degradation degradation and the PDE, together with its intuitive understanding as a transport equation and the implementation of different rates will be useful. One indicator for this is that the analytical solution of the ODE has been already used elsewhere [254].

Appendix A: ODE Model - Rhythmic Production and Rhythmic Degradation

All following derivations from this chapter plus the accompanying text were taken from the supplement of our published paper "Rhythmic Degradation Explains and Unifies Circadian Transcriptome and Proteome Data", S. Lück, K. Thurley, P.F. Thaben, P.O. Westermark in Cell Reports, 2014 [179]. The solution of the model derivation, its analytical and numerical justification and the error propagation used in the statistical test to test for rhythmic post-transcriptional regulation were my main contribution to this publication. The text has been slightly modified to fit to the text of this thesis.

A.1 ODE Model

A biomolecule x is rhythmically produced and rhythmically degraded:

$$\frac{dx}{dt} = u(t) - v(t)x(t), \tag{A.1}$$

where $u(t)$ and $v(t)$ are a zero-order-production coefficient (unit = abundance $\times$ time^{-1}) and a first-order-degradation coefficient (unit = time^{-1}), respectively. The degradation rate coefficient $u(t)$ and the half-life $t_{1/2}(t)$ are connected by the relationship

$$v(t) = \frac{\log 2}{t_{1/2}(t)}. \tag{A.2}$$

We use both terms interchangeably here, as well as in the main text. Equation A.1 is a linear differential equation with periodic coefficients. It has the ω-periodic solution [255]:

$$x(t) = x_0 e^{\int_0^t v(s)\mathrm{d}s} + \int_0^t u(t')\, e^{t' - \int_0^t v(s)\mathrm{d}s}\, \mathrm{d}t'. \tag{A.3}$$

However, even for simple cases, the integrations in Equation A.3 cannot be performed analytically.

In circadian biology, data (e.g. gene expression profiles) are often well described by cosine-

shaped functions, and we assume:

$$\begin{aligned} u(t) &= k(1 + A_{\text{prod}} \cos(\omega t - \phi_{\text{prod}})) \\ v(t) &= \gamma(1 + A_{\text{deg}} \cos(\omega t - \phi_{\text{deg}})). \end{aligned} \tag{A.4}$$

Thus, we assume the time-dependent production degradation rate coefficients $u(t)$ and $v(t)$ are cosine-shaped functions determined by their mean values k and γ, the relative amplitudes A_{prod} and A_{deg}, the angular frequency ω and the phases ϕ_{prod} and ϕ_{deg}. The relative amplitudes take values between 0 and 1 and are indicators for the strength of the oscillation. The relative amplitude is normally the amplitude parameter that can be best inferred from experimental data, as absolute concentrations are rarely determined. The angular frequency is set to

Without loss of generality, the phase ϕ_{deg} can be described with reference to the phase of production ϕ_{prod}, so that only the phase difference $\Delta\phi_{\text{deg}} = \phi_{\text{deg}} - \phi_{\text{prod}}$ between production and degradation needs to be considered. The model then takes the form

$$\begin{aligned} \frac{dx}{dt} &= u(t) - v(t)x(t) \\ u(t) &= k(1 + A_{\text{prod}} \cos(\omega t)) \\ v(t) &= \gamma(1 + A_{\text{deg}} \cos(\omega t - \Delta\phi_{\text{deg}})). \end{aligned} \tag{A.5}$$

The phase difference $\Delta\phi_{\text{deg}}$ takes values between $-\pi$ and π, i.e. between -12 h and $+12$ h in the circadian time frame. A negative value indicates that the degradation peaks before the production.

A.1.1 Constant Degradation: Exact Solution

In the following, the exact solution (Equation 2.4, 2.5 and 2.6 in Section 2.1 in the main text) for the special case of constant degradation ($A_{\text{deg}} = 0$) is derived. The solution of model Equation A.5 is found by standard techniques for solving ordinary differential equations [256]. After an initial transient (i.e., for $t \to \infty$), the abundance is described by a cosine-shaped function:

$$x(t) = M_x \left(1 + A_x \cos(\omega t - \phi_{\text{x}})\right), \tag{A.6}$$

with magnitude M_x, relative amplitude A_x, and phase ϕ_{x}. The abundance x is described with its phase shift with respect to the production rate $\Delta\phi_{\text{x}} = \phi_{\text{x}} - \phi_{\text{prod}}$:

$$x(t) = M_x \left(1 + A_x \cos(\omega t - \Delta\phi_{\text{x}})\right). \tag{A.7}$$

The solution for mean, amplitude and phase read then:

$$\begin{aligned} M_x &= \frac{k}{\gamma} \\ A_x &= \frac{\gamma A_{\text{prod}}}{\sqrt{\gamma^2+\omega^2}} \\ \phi_x &= \phi_{\text{prod}} + \arctan\left(\frac{\omega}{\gamma}\right) \quad \text{or} \\ \Delta\phi_{\text{x}} &= \arctan\left(\frac{\omega}{\gamma}\right). \end{aligned} \tag{A.8}$$

The last expressions show that in the case of constant degradation, the mean concentration equals the non-oscillatory steady-state (i.e., $A_{\text{prod}} = A_{\text{deg}} = 0$ and $dx/dt = 0$ in Equation A.5).

In Equations A.8, the range of $\Delta\phi_{\text{x}}$ is limited to the range of the arctan function for positive argument, which is between 0 and $\pi/2$, so that $\Delta\phi_{\text{x}}$ can only vary between 0 and 6 hours on the circadian time scale. Another direct conclusion from Equations A.8 is that $M \to \infty$ and $A_x \to 0$ for $\gamma \to 0$. This means that for long half-lives, magnitudes increase but amplitudes vanish, so that circadian rhythms are lost.

A.1.2 Rhythmic Degradation: Approximation with Fourier Expansion and Harmonic Balancing

Floquet theory guarantees that the solution to Equation A.1 is an ω-periodic function [255]. Thus, after an initial transient, the abundance x can be approximated by Fourier expansion to the order of n:

$$x(t) \approx A_0 + A_1\cos(\omega t) + B_1\sin(\omega t) + ... + A_n\cos(n\omega t) + B_n\sin(n\omega t). \tag{A.9}$$

The exact Fourier coefficients in such a truncated Fourier expansion cannot be calculated in closed form. Therefore, we use the following idea: We plug the truncated Fourier expansion Equation A.9 into the model, Equation A.1, and compare the coefficients. This leads to a system of linear equations which can be solved to obtain the approximate coefficients A_0, A_1, B_1, ..., A_n, B_n. These coefficients are not the exact Fourier coefficients, since we use a truncated Fourier series to obtain the describing linear equations. Below, we show that the error of this approximation vanishes as the expansion becomes longer, and that the coefficients then converge to the exact Fourier coefficients.

Numerical solutions generally exhibit shapes close to cosine-functions, this means that already a Fourier expansion to the first order is generally sufficient, see also numerical validation,

Section A.1.4, to describe the solution of the full model Equation A.5. Hence we obtain

$$\begin{aligned} A_0 =& \frac{k}{\gamma}\frac{2(\gamma^2+\omega^2)+A_{\text{prod}}A_{\text{deg}}\gamma(\omega\sin(\Delta\phi_{\text{deg}})-\gamma\cos(\Delta\phi_{\text{deg}}))}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} \\ A_1 =& \frac{2k\gamma(A_{\text{prod}}-A_{\text{deg}}\cos(\Delta\phi_{\text{deg}}))-2kA_{\text{deg}}\omega\sin(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} - \\ & \frac{A_{\text{prod}}A_{\text{deg}}^2k\gamma\sin^2(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} \\ B_1 =& \frac{2k\omega(A_{\text{prod}}-A_{\text{deg}}\cos(\Delta\phi_{\text{deg}}))+2kA_{\text{deg}}\gamma\sin(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} - \\ & \frac{A_{\text{prod}}A_{\text{deg}}^2k\gamma\sin(\Delta\phi_{\text{deg}})\cos(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2}\,. \end{aligned} \tag{A.10}$$

The relative amplitudes A_{prod} and A_{deg} take values smaller than or equal to 1, and the same is true for the absolute values $\sin(\Delta\phi_{\text{deg}})$ and $\cos(\Delta\phi_{\text{deg}})$. Mixed products of these terms as $A_{\text{prod}}A_{\text{deg}}^2\sin(\Delta\phi_{\text{deg}})\cos(\Delta\phi_{\text{deg}})$ and $A_{\text{prod}}A_{\text{deg}}^2\sin^2(\Delta\phi_{\text{deg}})$ are small, and are therefore neglected in the following, to obtain a more convenient approximation:

$$\begin{aligned} A_0 &= \frac{k}{\gamma}\frac{2(\gamma^2+\omega^2)+A_{\text{prod}}A_{\text{deg}}\gamma(\omega\sin(\Delta\phi_{\text{deg}})-\gamma\cos(\Delta\phi_{\text{deg}}))}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} \\ A_1 &\approx \frac{2k\gamma(A_{\text{prod}}-A_{\text{deg}}\cos(\Delta\phi_{\text{deg}}))-2kA_{\text{deg}}\omega\sin(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2} \\ B_1 &\approx \frac{2k\omega(A_{\text{prod}}-A_{\text{deg}}\cos(\Delta\phi_{\text{deg}}))+2kA_{\text{deg}}\gamma\sin(\Delta\phi_{\text{deg}})}{2(\gamma^2+\omega^2)-A_{\text{deg}}^2\gamma^2}. \end{aligned} \tag{A.11}$$

A Fourier Series to the first order describes a harmonic function as linear combination of a cosine function and a sine function. We reformulate this linear combination to gain a cosine function with a relative amplitude A_x and a phase $\Delta\phi_{\text{x}}$ [256] (cp. also Equation A.7):

$$x(t) = A_0 + A_1\cos(\omega t) + B_1\sin(\omega t) = M_x(1 + A_x\cos(\omega t - \Delta\phi_{\text{x}})), \tag{A.12}$$

where

$$\begin{aligned} M_x =& A_0 \\ A_x =& \frac{1}{A_0}\sqrt{A_1^2+B_1^2} \\ \Delta\phi_{\text{x}} =& \arctan 2\,(B_1, A_1)\,. \end{aligned} \tag{A.13}$$

The function $\arctan 2(y,x)$ is the arctangent function with two arguments, which computes the principal value of the argument function applied to the complex number $x+iy$. The definition of $\arctan 2(y,x)$ is given in Figure A.1. With this ansatz we neglect higher order terms in the Fourier expansion, which is justified below.

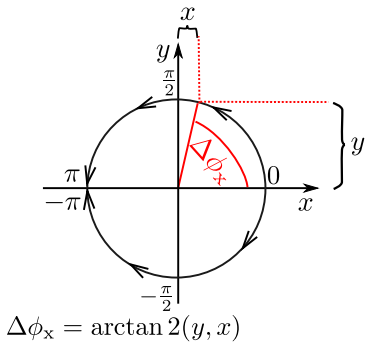


Figure A.1: Definition of the function $\arctan 2(y, x)$. The function $\arctan(y/x)$ covers a range from $-\pi/2$ to $+\pi/2$, which would not be suitable to describe an arbitrary phase shift. To consider all possible phases, we use the function $\arctan 2(y, x)$, which covers the full range from $-\pi$ to $+\pi$ and is defined as follows: $\Delta\phi_x = \arctan 2(y, x)$ is the angle between (x, y) and $(y = 0, x > 0)$ in the x-y plane. If $y > 0$, the angle is taken counterclockwise from 0 to π, and if $y < 0$, then the angle is taken clockwise from 0 to $-\pi$.

With Equations A.11 we obtain the following expressions:

$$\begin{aligned} M_x =& \frac{k}{\gamma} \frac{2(\gamma^2 + \omega^2 - \frac{A_{\text{prod}} A_{\text{deg}}}{2} \gamma \left(\omega \sin(\Delta\phi_{\text{deg}}) + \gamma \cos(\Delta\phi_{\text{deg}})\right))}{2(\gamma^2 + \omega^2) - \gamma^2 A_{\text{deg}}^2} \\ A_x =& \frac{\gamma \sqrt{A_{\text{prod}}^2 + A_{\text{deg}}^2 - 2 A_{\text{prod}} A_{\text{deg}} \cos(\Delta\phi_{\text{deg}})}}{\sqrt{\gamma^2 + \omega^2} - \frac{A_{\text{prod}} A_{\text{deg}}}{2} \gamma \left(\omega \sin(\Delta\phi_{\text{deg}}) + \gamma \cos(\Delta\phi_{\text{deg}})\right)} \\ \Delta\phi_{\text{x}} =& \arctan 2 \begin{pmatrix} \omega \left(A_{\text{prod}} - A_{\text{deg}} \cos(\Delta\phi_{\text{deg}})\right) - A_{\text{deg}} \gamma \sin(\Delta\phi_{\text{deg}}), \\ \gamma \left(A_{\text{prod}} - A_{\text{deg}} \cos(\Delta\phi_{\text{deg}})\right) + A_{\text{deg}} \omega \sin(\Delta\phi_{\text{deg}}) \end{pmatrix} . \end{aligned} \tag{A.14}$$

For the special case of constant degradation ($A_{\text{deg}} = 0$), these equations reduce to the exact expressions, Equations A.8. **As defined previously, we used the relative phase $\Delta\phi_{\mathbf{deg}} = \phi_{\mathbf{deg}} - \phi_{\mathbf{prod}}$.**

Vector Representation of Phase and Amplitude

The Fourier expansion resulted in a good approximation for biologically relevant parameter values, see also numerical validation, Section A.1.4, and if needed it is possible to use the same approach and expand to a higher order to achieve higher accuracy. However, Equations A.14 do not provide intuitive insight into the properties of phase and amplitude of a rhythmically degraded biomolecule. **In the main text (see also Section 2.5), we introduce phase and relative amplitude as a result of a vector addition in the complex plane (Equations 2.10, 2.11. In the following we derive this description from the Fourier expansion derived above.**

The phase $\Delta\phi_{\text{x}}$ is formulated in Equations A.14 with an $\arctan 2(y, x)$ function. This equation can be reinterpreted as a calculation in a two-dimensional vector space. Then, the phase is the

angle of a vector with the entries x and y:

$$\Delta\phi_{\mathrm{x}} = \arctan 2(y,x) \stackrel{\mathrm{def}}{=} \mathrm{angle}\begin{pmatrix} x \\ y \end{pmatrix} .$$

Now the phase is described as the result of a vector calculation. The expression in Equation A.14 can be rewritten as an angle:

$$\Delta\phi_{\mathrm{x}} = \mathrm{angle}\left\{ \left[A_{\mathrm{prod}} - A_{\mathrm{deg}}\mathbf{R}(\Delta\phi_{\mathrm{x}}) \right] \begin{pmatrix} \gamma \\ \omega \end{pmatrix} \right\} . \tag{A.15}$$

Here, $\mathbf{R}(\Delta\phi_{\mathrm{x}})$ is a rotation matrix that describes a counterclockwise rotation by the angle $\Delta\phi_{\mathrm{x}}$:

$$\mathbf{R} = \begin{pmatrix} \cos(\Delta\phi_{\mathrm{x}}) & -\sin(\Delta\phi_{\mathrm{x}}) \\ \sin(\Delta\phi_{\mathrm{x}}) & \cos(\Delta\phi_{\mathrm{x}}) \end{pmatrix} .$$

From the two-dimensional vector space to the complex plane, a rotation matrix by the angle ϕ translates to $e^{i\phi}$ and a vector with positive entries x and y to $\sqrt{x^2+y^2}e^{i\arctan(y/x)}$. Thus, Equation A.15 translates to:

$$\begin{aligned} \Delta\phi_{\mathrm{x}} &= \arg\left\{ \left[A_{\mathrm{prod}} - A_{\mathrm{deg}}e^{i\Delta\phi_{\mathrm{deg}}} \right] \sqrt{\gamma^2+\omega^2}e^{i\arctan(\omega/\gamma)} \right\} \\ \Delta\phi_{\mathrm{x}} &= \arg\left(A_{\mathrm{prod}} - A_{\mathrm{deg}}e^{i\Delta\phi_{\mathrm{deg}}} \right) + \arg\left(e^{i\arctan(\omega/\gamma)} \right) . \end{aligned} \tag{A.16}$$

To derive the Fourier approximation we reduced the number of parameters by describing all phases in relation to the phase of the production. We now return to a description in the absolute time frame and rewrite $\Delta\phi_{\mathrm{deg}} = \phi_{\mathrm{deg}} - \phi_{\mathrm{prod}}$ and $\Delta\phi_{\mathrm{x}} = \phi_x - \phi_{\mathrm{prod}}$. This gives for the calculation A.16 in the complex space:

$$\phi_x = \arg\left(A_{\mathrm{prod}}e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{deg}}e^{i\phi_{\mathrm{deg}}} \right) + \arctan\left(\frac{\omega}{\gamma}\right) . \tag{A.17}$$

This is the vector representation of the phase in the main text (Equation 2.11). We introduced the term $A_{\mathrm{prod}}e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{deg}}e^{i\phi_{\mathrm{deg}}}$ in the main text as the so-called "production-degradation vector". It describes the influence of only the oscillatory part of the production and degradation rate independent from their mean rates. The absolute value of the effective production is:

$$\left| A_{\mathrm{prod}}e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{deg}}e^{i\phi_{\mathrm{deg}}} \right| = \sqrt{A_{\mathrm{prod}}^2 + A_{\mathrm{deg}}^2 - 2A_{\mathrm{prod}}A_{\mathrm{deg}}\cos(\phi_{\mathrm{deg}} - \phi_{\mathrm{prod}})} . \tag{A.18}$$

Replacing $\Delta\phi_{\mathrm{deg}} = \phi_{\mathrm{deg}} - \phi_{\mathrm{prod}}$, the square root in Equation A.18 can be identified in the previously derived amplitude of x (see Equation A.14), which yields:

$$A_x = \frac{\gamma \left| A_{\mathrm{prod}}e^{i\phi_{\mathrm{prod}}} - A_{\mathrm{deg}}e^{i\phi_{\mathrm{deg}}} \right|}{\sqrt{\gamma^2+\omega^2} - \frac{A_{\mathrm{prod}}A_{\mathrm{deg}}}{2}\gamma\left(\omega\sin(\phi_{\mathrm{deg}} - \phi_{\mathrm{prod}}) + \gamma\cos(\phi_{\mathrm{deg}} - \phi_{\mathrm{prod}})\right)} . \tag{A.19}$$

We refer to the factor $\frac{A_{\text{prod}}A_{\text{deg}}}{2}\gamma\left(\omega\sin(\phi_{\text{deg}}-\phi_{\text{prod}})+\gamma\cos(\phi_{\text{deg}}-\phi_{\text{prod}})\right)$ as a correction factor and replace it by C in the main text. **With this step we have derived the vector description of the amplitude (Equation 2.10 in the main text).**

A.1.3 Analytical Validation of the Approximation

The approximation Equation A.14 is derived by a truncated Fourier expansion. In the following we justify this truncation and the convergence of the approximation. Assume that $u(t)$ and $v(t)$ in Equation A.1 are smooth and bounded ω-periodic functions. This ensures that $x(t)$ and $x'(t)$ are smooth and bounded ω-periodic functions [255], and therefore the Fourier series of $x(t)$ converges [257]. Moreover, the sequence of the Fourier coefficients $A_0, A_1, \cdots$ and $B_1, B_2, \cdots$ decays in a way that $|A_n| < K/n^2$, $|B_n| < K/n^2$, with a constant K independent of n. If in addition $u'(t)$ and $v'(t)$ are smooth functions, we can take the derivative on the right-hand side of Equation A.1, and therefore also on the left-hand side, and so on. If the differential coefficients of a function $x(t)$ are bounded and continuous up to the order $p-1$, the sequence decays even faster [257]:

$$|A_n| < K/n^p, \quad |B_n| < K/n^p. \tag{A.20}$$

Therefore, the smoothness of the input functions $u(t)$ and $v(t)$ determines the error made by truncating the Fourier series of the (unknown) function $x(t)$. Since cosine functions, which we use here (Equations A.5), are indefinitely often differentiable, a reasonably good approximation can be expected, even if only the first Fourier coefficients are taken into account. We were able to justify this expectation by numeric simulations (see Figure A.4). In principle, the approximation procedure outlined in the last section can also be used for other types of input functions $u(t)$ and $v(t)$ such as square-waves or spikes. However, the reasoning above suggests that the convergence of the Fourier series will be weaker for input functions that are not as smooth as cosine functions. Thus, it might be necessary to include higher order Fourier coefficients in that case.

There is no exact formula for the Fourier coefficients of $x(t)$, which is why we derived an approximation procedure for the first Fourier coefficients (Equation A.10). Here, we show that such approximations improves with the number of terms of the series that are included. A Fourier expansion to the nth term results in the following set of linear equations:

$$\begin{pmatrix} \gamma & \mathbf{A_{1\times 2}} & & & \cdots & & & \\ \mathbf{A_{2\times 1}} & \mathbf{RS_1} & \mathbf{RS^-} & & \cdots & & 0 & \\ & \mathbf{RS^+} & \mathbf{RS_2} & \mathbf{RS^-} & \cdots & & & \\ & & \mathbf{RS^+} & \mathbf{RS_3} & \cdots & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ & 0 & & & \cdots & \mathbf{RS^+} & \mathbf{RS_{n-1}} & \mathbf{RS^-} \\ & & & & \cdots & & \mathbf{RS^+} & \mathbf{RS_n} \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ \mathbf{C_1} \\ \mathbf{C_2} \\ \mathbf{C_3} \\ \vdots \\ \mathbf{C_{n-1}} \\ \mathbf{C_n} \end{pmatrix} = \begin{pmatrix} a \\ \mathbf{vec} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{A.21}$$

with these abbreviations:

$$\mathbf{A_{1\times 2}} = \begin{pmatrix} \frac{\gamma}{2} A_{\text{deg}} \cos(\Delta\phi_{\text{deg}}) & \frac{\gamma}{2} A_{\text{deg}} \sin(\Delta\phi_{\text{deg}}) \end{pmatrix} , \quad \mathbf{A_{2\times 1}} = \begin{pmatrix} \gamma A_{\text{deg}} \cos(\Delta\phi_{\text{deg}}) \\ \gamma A_{\text{deg}} \sin(\Delta\phi_{\text{deg}}) \end{pmatrix} ,$$

$$\mathbf{RS}^{-} = \frac{\gamma}{2} A_{\text{deg}} \begin{pmatrix} \cos(\Delta\phi_{\text{deg}}) & \sin(\Delta\phi_{\text{deg}}) \\ -\sin(\Delta\phi_{\text{deg}}) & \cos(\Delta\phi_{\text{deg}}) \end{pmatrix} , \quad \mathbf{RS}^{+} = \frac{\gamma}{2} A_{\text{deg}} \begin{pmatrix} \cos(\Delta\phi_{\text{deg}}) & -\sin(\Delta\phi_{\text{deg}}) \\ \sin(\Delta\phi_{\text{deg}}) & \cos(\Delta\phi_{\text{deg}}) \end{pmatrix} ,$$

$$\mathbf{RS_n} = \begin{pmatrix} \gamma & n\omega \\ -n\omega & \gamma \end{pmatrix} , \ \mathbf{C_n} = \begin{pmatrix} A_n \\ B_n \end{pmatrix} \quad \text{and} \quad \mathbf{vec} = \begin{pmatrix} A_{\text{prod}} k \\ 0 \end{pmatrix} .$$

The matrix that defines the approximated Fourier coefficients has a block-structure and by following this pattern, the matrix can readily be extended to higher orders. It consists of mainly three submatrices $\mathbf{RS}^{+}$, $\mathbf{RS}^{-}$ and $\mathbf{RS_n}$. The first two submatrices can be identified as matrices that rotate either clockwise ($\mathbf{RS}^{+}$) or counterclockwise ($\mathbf{RS}^{-}$) by the angle $\Delta\phi_{\text{deg}}$, and scale by the constant factor $\frac{\gamma}{2} A_{\text{deg}}$. The third submatrix $\mathbf{RS_n}$ can also be interpreted as a rotation-scaling-matrix:

$$\begin{aligned} \mathbf{RS_n} &= \sqrt{\gamma^2 + n^2\omega^2} \begin{pmatrix} \frac{\gamma}{\sqrt{\gamma^2+n^2\omega^2}} & \frac{n\omega}{\sqrt{\gamma^2+n^2\omega^2}} \\ \frac{-n\omega}{\sqrt{\gamma^2+n^2\omega^2}} & \frac{\gamma}{\sqrt{\gamma^2+n^2\omega^2}} \end{pmatrix} = \sqrt{\gamma^2 + n^2\omega^2} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} , \\ \mathbf{RS_n} &= \sqrt{\gamma^2 + n^2\omega^2}\, \mathbf{R}(\theta) . \end{aligned} \tag{A.22}$$

It describes a rotation $\mathbf{R}(\theta)$ by the angle $\theta = \arctan(n\omega/\gamma)$ and scaling by the factor $\sqrt{\gamma^2 + n^2\omega^2}$. With increasing order n this scaling factor becomes larger. In the following, we justify that in a truncated Fourier expansion to the order of n, the coefficients have a small approximation error compared to the exact Fourier coefficients and that with increasing n, the approximated Fourier coefficients converge to the exact Fourier coefficients.

The block-structure of the matrix shows that the nth coefficients are only determined by the neighboring coefficients $\mathbf{C_{n+1}}$ and $\mathbf{C_{n-1}}$:

$$\mathbf{RS_n} \cdot \begin{pmatrix} A_n \\ B_n \end{pmatrix} = \mathbf{RS}^{-} \cdot \begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} + \mathbf{RS}^{+} \cdot \begin{pmatrix} A_{n-1} \\ B_{n-1} \end{pmatrix} . \tag{A.23}$$

We reformulate Equation A.23 by multiplying with the inverse of $\mathbf{RS_n}$. Using Equation A.22 yields:

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = \frac{1}{\sqrt{\gamma^2 + n^2\omega^2}}\, \mathbf{R}(-\theta)\mathbf{RS}^{-} \cdot \begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} + \frac{1}{\sqrt{\gamma^2 + n^2\omega^2}}\, \mathbf{R}(-\theta)\mathbf{RS}^{+} \cdot \begin{pmatrix} A_{n-1} \\ B_{n-1} \end{pmatrix} . \tag{A.24}$$

If the coefficients A_{n+1} and B_{n+1} were known and identical to the exact Fourier coefficients, then with Equation A.24 and with the equations for the following coefficients defined by the block-structure of the matrix, every exact Fourier coefficient of order n and lower could be solved

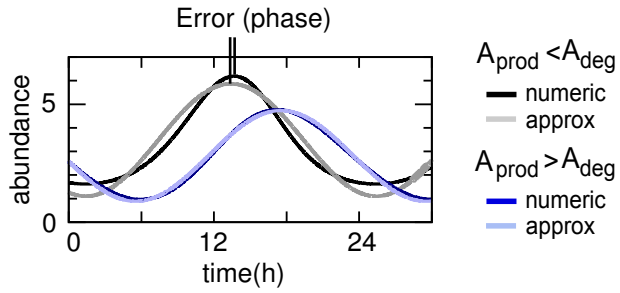


Figure A.2: Numerical simulation and analytical approximation of the model. Marked is the error made by the approximation. Parameter values: $A_{\text{prod}} = 0.2$ and $A_{\text{deg}} = 0.8$ (black/gray) or $A_{\text{prod}} = 0.8$ and $A_{\text{deg}} = 0.2$ (blue), $\omega = 2\pi/24$ h, $\gamma = 0.35$ h^{-1}, $k = 1$ h^{-1}, $\Delta\phi_{\text{deg}} = \pi/2$.

for. However, the coefficients A_{n+1} and B_{n+1} are always unknown, and in our approximation the Fourier series are truncated by setting the term $\begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix}$ to zero:

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} \approx \frac{1}{\sqrt{\gamma^2 + n^2\omega^2}} \, \mathbf{R}(-\theta)\mathbf{R}\mathbf{S}^+ \cdot \begin{pmatrix} A_{n-1} \\ B_{n-1} \end{pmatrix}. \tag{A.25}$$

The coefficients $\begin{pmatrix} A_i \\ B_i \end{pmatrix}$, $i < n$, are also defined only by their neighboring coefficients, higher coefficients do not contribute. This means setting the $(n+1)$th coefficients to zero is the only approximation we apply to derive the coefficients. We define a relative error as the ratio of the approximation error to the nth coefficients:

$$\text{error} \stackrel{\text{def}}{=} \frac{\left| (\gamma^2 + n^2\omega^2)^{-\frac{1}{2}} \, \mathbf{R}(-\theta)\mathbf{R}\mathbf{S}^- \cdot \begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} \right|}{\left| \begin{pmatrix} A_n \\ B_n \end{pmatrix} \right|} = \frac{\gamma A_{\text{deg}}}{2\sqrt{\gamma^2 + n^2\omega^2}} \frac{\left| \begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} \right|}{\left| \begin{pmatrix} A_n \\ B_n \end{pmatrix} \right|}. \tag{A.26}$$

Here, $|\cdot|$ denotes the Euclidean vector norm. With increasing n the ratio in front of the vector norm ratio becomes small. In addition to that, the vector norm ratio becomes small since Fourier coefficients decay fast with increasing order, as outlined above. Thus, with increasing order n, the relative approximation error becomes small and the coefficients of the truncated Fourier series converge to the exact Fourier coefficients.

A.1.4 Numerical Validation of the Approximation

To validate the Fourier approximation, we compared the approximation to numerically computed abundances for different parameter sets by variation of the mean degradation rate γ, the phase difference between production and degradation $\Delta\phi_{\text{deg}}$,

and the relative amplitudes $A_{\mathbf{prod}}$ and $A_{\mathbf{deg}}$, as outlined in the following. We denote the difference between numerical solution and approximation by the model as "Error", see also Fig. A.2. For example, the Error$(\phi_{\text{prod}} - \phi_x)$ stands for:

$$\text{Error}(\phi_{\text{prod}} - \phi_x) = (\phi_{\text{prod}} - \phi_x)_{\text{numerical}} - (\phi_{\text{prod}} - \phi_x)_{\text{approximation}} \tag{A.27}$$

The production term and degradation rate were either cosine-functions simulated with

$$u(t) = k\,(1 + A_{\text{prod}} \cos(\omega t)) \quad \text{and} \tag{A.28}$$

$$v(t) = \gamma\,(1 + A_{\text{deg}} \cos(\omega t - \Delta\phi_{\text{deg}}))\,, \tag{A.29}$$

or a Fourier series to the seventh order of periodic parabolas:

$$u(t) = \gamma\left(1 - \frac{A_{\text{prod}}}{6} + \frac{4A_{\text{prod}}}{\pi^2}\left(\cos(\omega t) + \frac{1}{2^2}\cos(2\omega t) + \frac{1}{3^2}\cos(3\omega t) + \frac{1}{4^2}\cos(4\omega t) + \frac{1}{5^2}\cos(5\omega t) + \frac{1}{6^2}\cos(6\omega t)\right)\right) \quad \text{and} \tag{A.30}$$

$$v(t) = \gamma\left(1 - \frac{A_{\text{deg}}}{6} + \frac{4A_{\text{deg}}}{\pi^2}\left(\cos(\omega t - \Delta\phi_{\text{deg}}) + \frac{1}{2^2}\cos(2(\omega t - \Delta\phi_{\text{deg}})) + \frac{1}{3^2}\cos(3(\omega t - \Delta\phi_{\text{deg}})) + \frac{1}{4^2}\cos(4(\omega t - \Delta\phi_{\text{deg}})) + \frac{1}{5^2}\cos(5(\omega t - \Delta\phi_{\text{deg}})) + \frac{1}{6^2}\cos(6(\omega t - \Delta\phi_{\text{deg}}))\right)\right). \tag{A.31}$$

As an example we plot a degradation rate defined by this Fourier series in Figure A.3. It reflects non-cosine-shaped time series similar to typical transcript abundances [5].

The difference between the numerical solution and approximation depends on the set of parameters and of the form of production and degradation rate. From the theoretical point of view we expect for both rate forms that the approximation by the model is close to the numerical solution, since the proposed rates are both smooth functions and therefore higher Fourier coefficients, which we do not include in the model, vanish fast, as we discussed earlier. However, for rates described by a Fourier series with higher terms, we expect less conformity.

For numerical simulation a C-script was used. To detect phase, amplitude and magnitude in the time series I used a similar algorithm as I later did to detect oscillation properties in the Goodwin model described in Appendix E.1. The same C-script was also used to produce Figure 2.9A.

In Figure A.4, we show the absolute error between model prediction and numerical solution of cosine-shaped rates, in Figure A.5 the absolute error for rates defined by the Fourier series of periodic parabolas. As expected for the rates modeled by periodic parabolas, the absolute error is higher than for cosine-shaped rates although there is no qualitative change. **For most parameter combinations, the error is below 10% for both phase difference and amplitude.** The approximation is particularly accurate if the degradation rate oscillates weaker than the production term ($A_{\text{prod}} > A_{\text{deg}}$), and for small phase differences $\Delta\phi_{\text{deg}} \approx 0$.

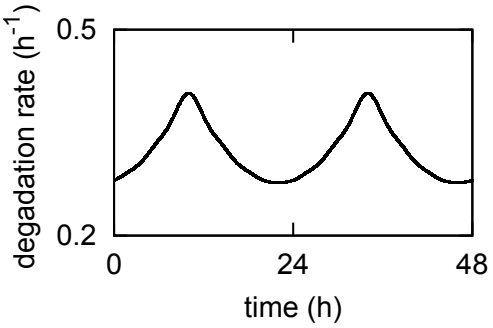


Figure A.3: Function with non-sinusoidal shape as in Equation A.31. Parameters: $\gamma = 0.35$ h^{-1}, $A_{\mathrm{deg}} = 0.2$, $\Delta\phi_{\mathrm{deg}} = 0.5\pi$.

Unknown Degradation Phase

In the main text we used our model to predict the unknown degradation rates for given production rates and RNA abundances (see Section 2.5.3). In the Section 2.5.3, we present equations for phase ϕ_{deg} and amplitude A_{deg} of the degradation (Equations 2.12, 2.13). Figures A.6 and A.7 show that these approximations of the phase and amplitude are particularly good at moderate to long half-lives. We compared the approximations to numerical simulations as follows: We first chose a certain parameter set [k, γ, A_{prod}, A_{deg}, ϕ_{prod}, ϕ_{deg}] and ran the numerical integration of our model Equation A.5. Then, we estimated the relative amplitude A_x and phase ϕ_{x} from the obtained time series and used these parameters together with our approximation, Equation 7, to recalculate the values of A_{deg} and ϕ_{deg}. The differences between the values of A_{deg} and ϕ_{deg} in the parameter set and those obtained from the approximation show the accuracy of the latter.

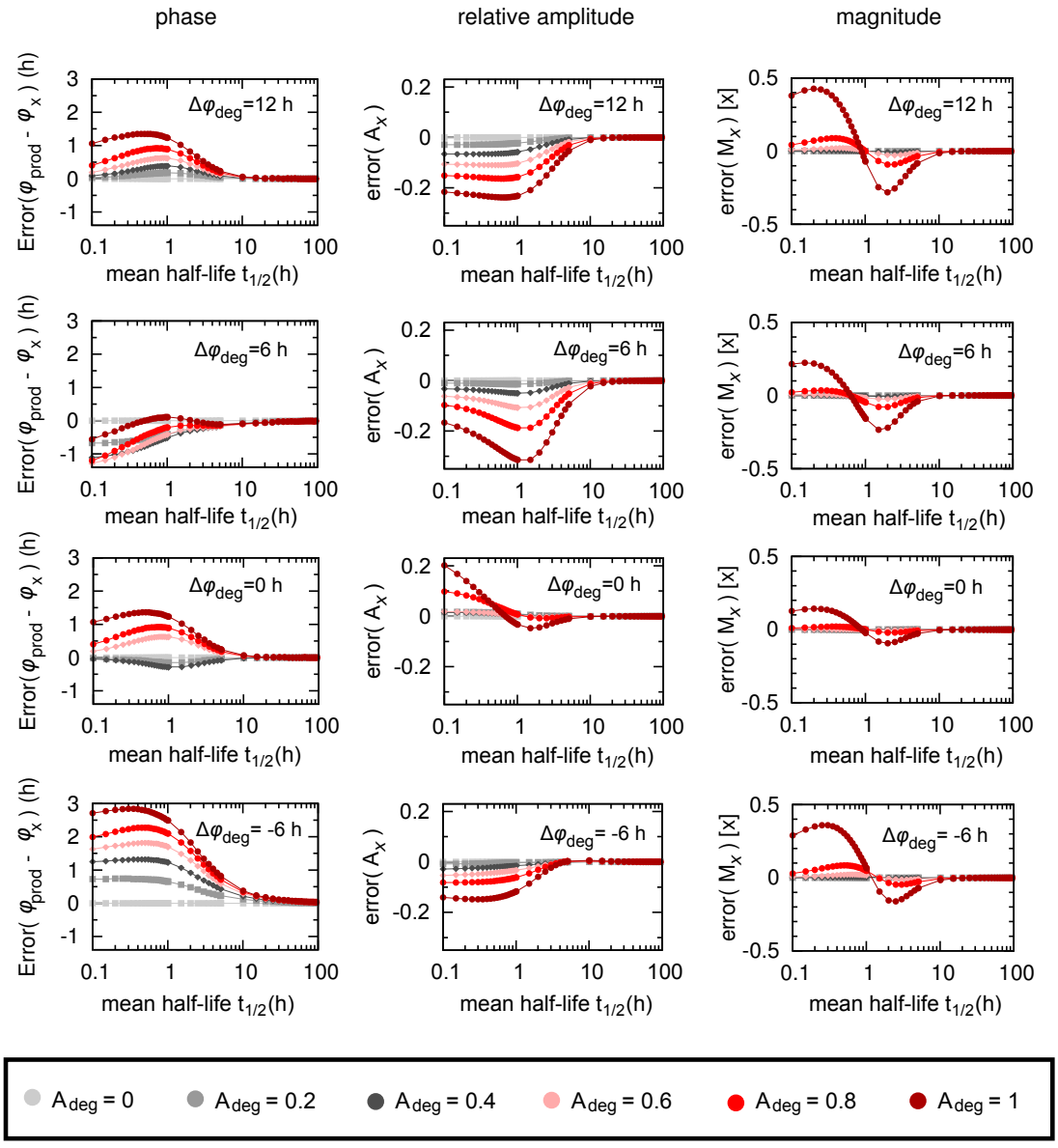


Figure A.4: Validation of the Fourier approximation. The production term and the degradation rate were modeled as cosine functions. The expressions for amplitude, phase and magnitude (Equations A.14) were compared with the numerical solution of Equations A.5 (see Experimental Procedures), shown are absolute errors. Parameters: $A_{\mathrm{prod}} = 0.5$, $k = 1\ \mathrm{h}^{-1}$.

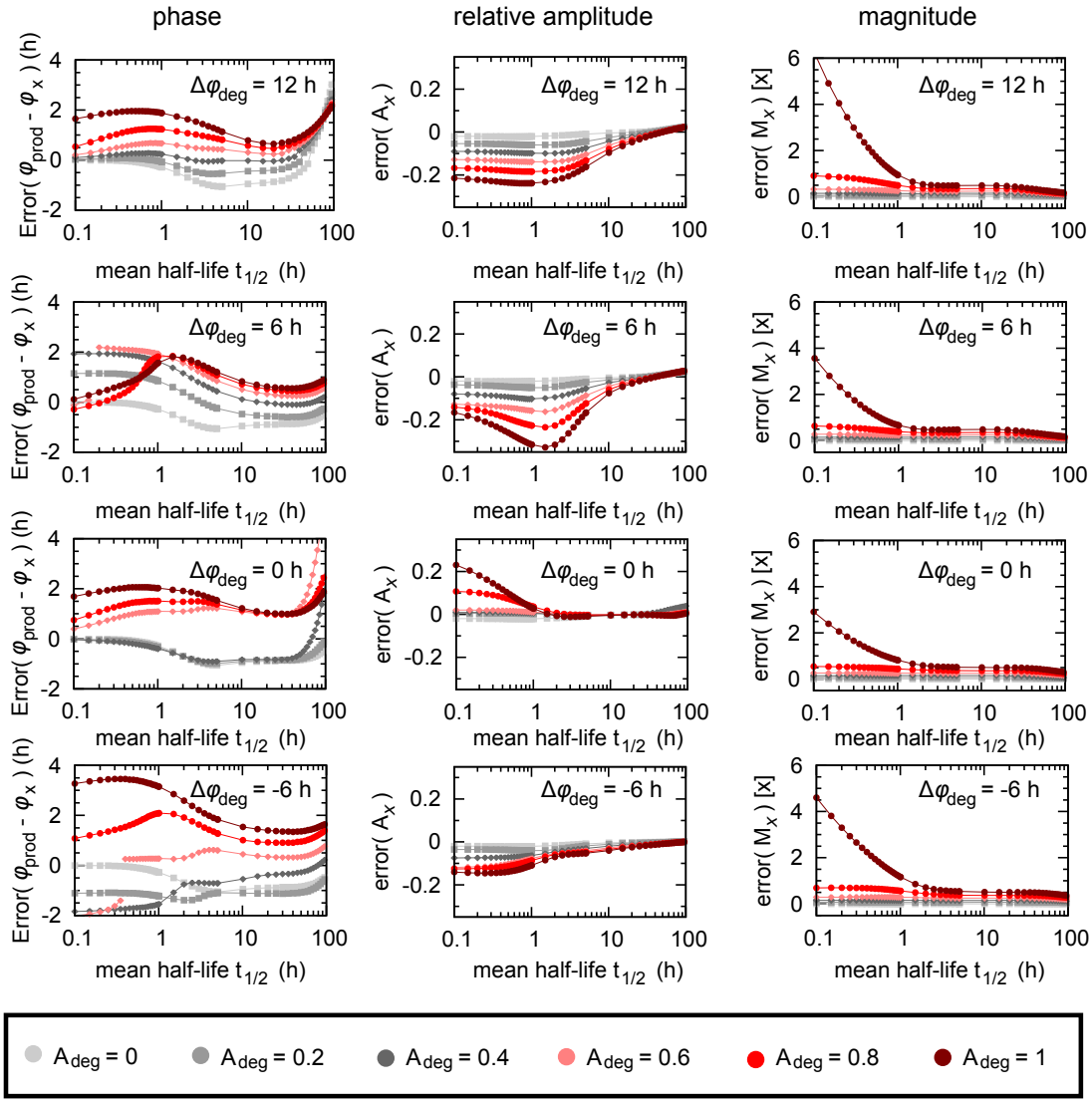


Figure A.5: Validation of the Fourier approximation. The production term and the degradation rate are modeled as non-sinoidal functions (see text). Same procedure and parameter choice as in Figure A.4.

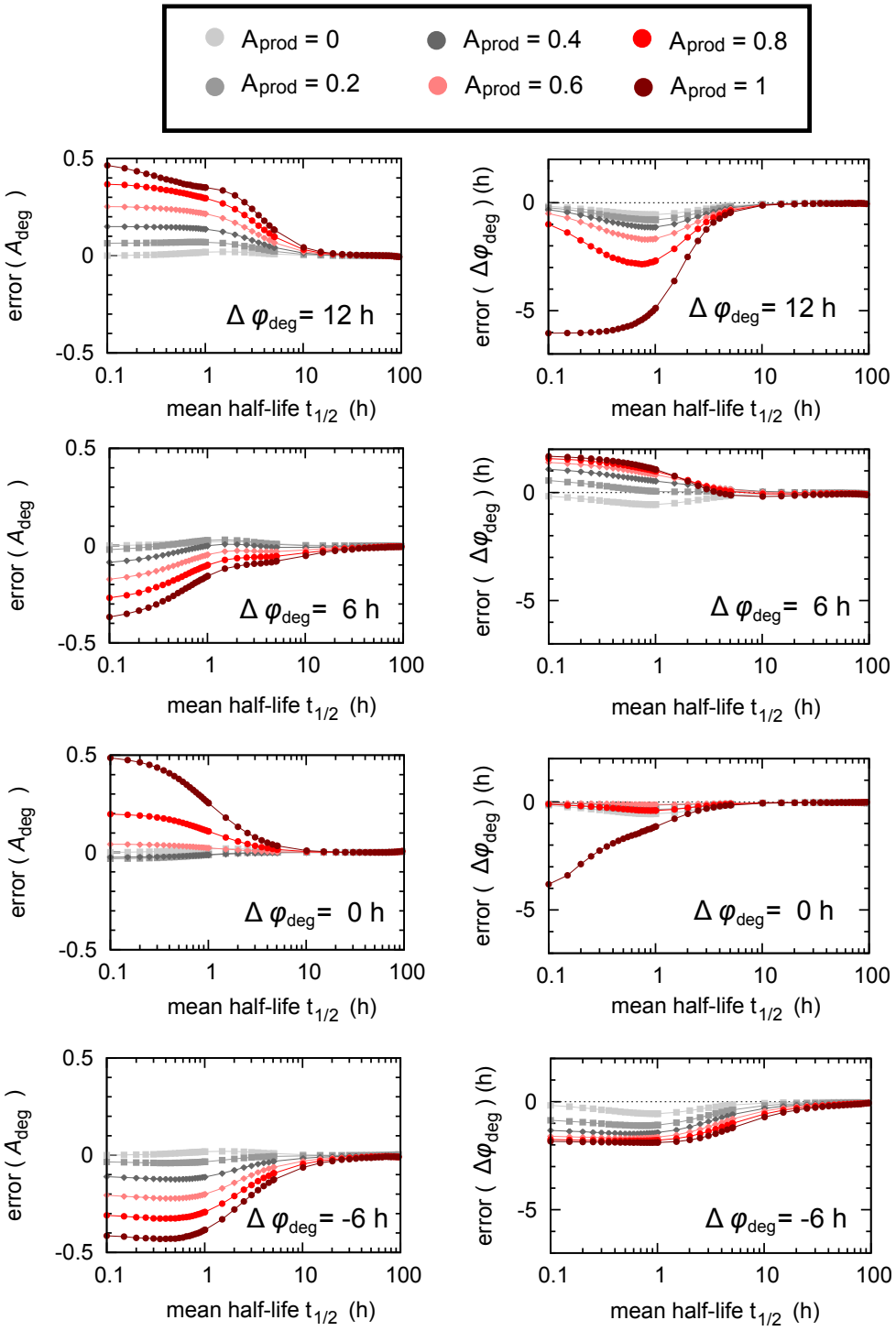


Figure A.6: Validation of the Fourier approximation for an unknown degradation rhythm at varying half-lives. The approximation (Equations 2.12, 2.13) is compared to the numerical solution of the inverse problem (see text). We show absolute errors at indicated parameter values. Parameters: $k = 1$, $A_{\text{deg}} = 0.5$

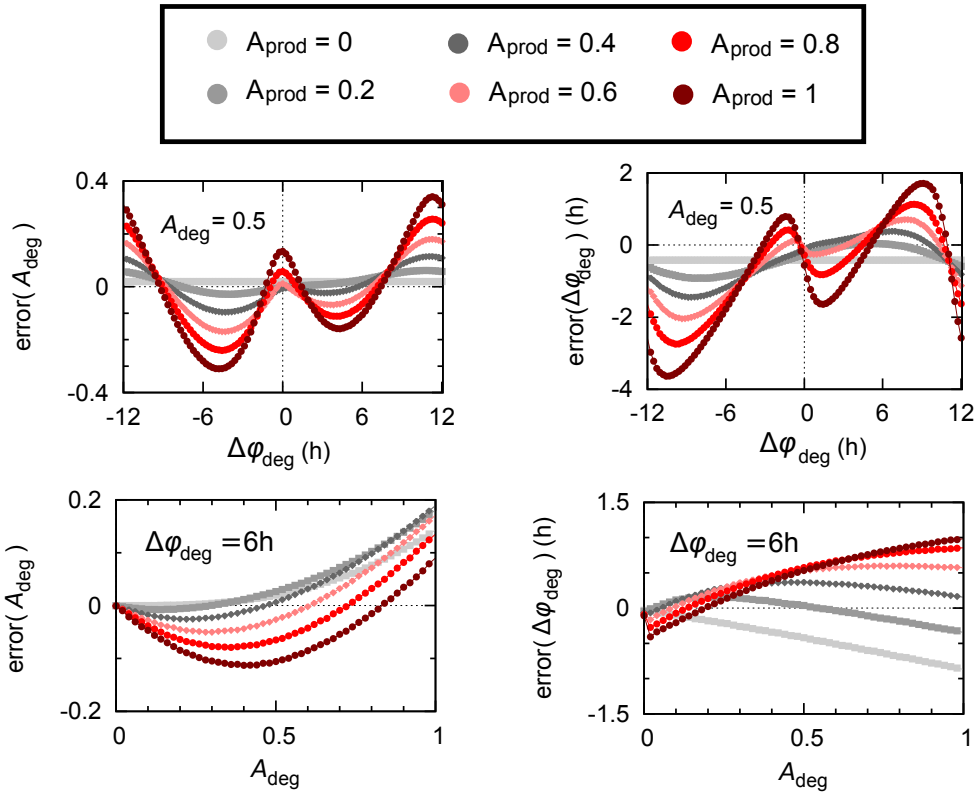


Figure A.7: Validation of the Fourier approximation for an unknown degradation rhythm at varying phases and relative amplitudes. $\gamma = 0.35\text{h}^{-1}$.

A.2 Additional Results from the ODE model

A.2.1 Error Propagation of Half-Lives for the Production-Degradation Vector

Here we derive the point estimate and the covariance matrix for the production-degradation vector. These estimates are needed for the chi-squared test described in Section 2.4.1 and implemented in the R package "patest".

We estimate amplitudes and phases using harmonic regression: Ordinary least squares fitting to a linear combination of cosine- and sine-functions:

$$x(t) = m + a\cos(\omega t) + b\sin(\omega t) + \epsilon, \tag{A.32}$$

where $x(t)$ stands for the abundance of a molecule at time t, and ϵ is an error term. The parameters a and b define (span) the circadian phase plane and constitute the vector x

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix},$$

and the fitting procedure results in estimates of a and b denoted $\boldsymbol{\mu_x}$ with symmetric covariance matrix $\mathbf{cov}_x$:

$$\boldsymbol{\mu}_{\mathbf{x}} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \tag{A.33}$$

$$\mathbf{cov}_{\mathbf{x}} = \begin{pmatrix} \sigma_a^2 & \sigma_{a,b} \\ \sigma_{a,b} & \sigma_b^2 \end{pmatrix}. \tag{A.34}$$

If the measured time series is evenly sampled and consists only of fully measured periods the covariance matrix reduces to a diagonal matrix:

$$\mathbf{cov}_{\mathbf{x}} = \begin{pmatrix} \sigma_{ab}^2 & 0 \\ 0 & \sigma_{ab}^2 \end{pmatrix}. \tag{A.35}$$

Estimates $\boldsymbol{\mu}_{\mathbf{x}}$ and $\mathbf{cov}_{\mathbf{x}}$ are obtained from the residuals of the fit [215, 258]. In the same way, a vector $\mathbf{p}$ representing production phase and amplitude is obtained from measurements of transcriptional activity for the case of mRNAs (in practice, these are often also abundances, for example of nascent mRNA), together with an estimated covariance matrix $\mathbf{cov}_{\mathbf{p}}$.

In our model for rhythmic degradation we define the so called production-degradation vector $\mathbf{pd}$, which is determined by a shift in the x- and y-coordinate of $\mathbf{x}$ depending on the half-life τ. Our chi-squared test is based on the difference between the production-degradation vector and the measured production vector. The null distribution for this difference can be estimated if an estimation for the covariance matrix of the production-degradation vector is available. Such an estimate can be obtained from the abundance vector (harmonic regression estimates from mature mRNA-seq data in the present study), the production vector (harmonic regression estimates from nascent mRNA-seq data), and the estimated covariance matrices for these. Also needed is the estimated half-life of the mRNA molecule, and the variance of this estimate. The

half-life τ of the molecule is measured with mean μ_τ and standard deviation σ_τ.

The x-component of the production-degradation vector **pd** follows:

$$\begin{aligned} x_{\mathrm{pd}} &= \frac{\sqrt{\gamma^2+\omega^2}}{\gamma}\sqrt{a^2+b^2}\cos\left(\operatorname{arctan2}(b,a)-\arctan(\omega/\gamma)\right) \\ &= \frac{\sqrt{\gamma^2+\omega^2}}{\gamma}\left(a\cos\left(\arctan(\omega/\gamma)\right)+b\sin\left(\arctan(\omega/\gamma)\right)\right) \\ &= \frac{(a\gamma+b\omega)\sqrt{1+\frac{\omega^2}{\gamma^2}}}{\sqrt{\gamma^2+\omega^2}} \\ &= a+\frac{b\omega}{\gamma}. \end{aligned} \tag{A.36}$$

Here, γ is the mean degradation rate which is inversely related to the half-life: $\gamma = \log(2)/\tau$. The y-component of the vector **pd** follows:

$$y_{\mathrm{pd}} = b - \frac{a\omega}{\gamma}. \tag{A.37}$$

We search the estimate $\boldsymbol{\mu}_{\mathbf{pd}}$ and covariance matrix $\mathbf{cov}_{\mathbf{pd}}$ for the vector **pd**:

$$\boldsymbol{\mu}_{\mathbf{pd}} = \begin{pmatrix} \mu_{x_{pd}} \\ \mu_{y_{pd}} \end{pmatrix}, \tag{A.38}$$

$$\mathbf{cov}_{\mathbf{pd}} = \begin{pmatrix} \sigma^2_{x_{\mathrm{pd}}} & \sigma_{x_{\mathrm{pd}},y_{\mathrm{pd}}} \\ \sigma_{x_{\mathrm{pd}},y_{\mathrm{pd}}} & \sigma^2_{y_{\mathrm{pd}}} \end{pmatrix}. \tag{A.39}$$

The components of the vector **pd** are functions of three variables, a, b, and τ. The mean $\boldsymbol{\mu}_{\mathbf{pd}}$ is then obtained from the definitions of **pd**, Eqns. (A.36 and A.37)

$$\boldsymbol{\mu}_{\mathbf{pd}} = \begin{pmatrix} \mu_a + \frac{\mu_b\mu_\tau}{\log(2)} \\ \mu_b - \frac{\mu_a\omega\mu_\tau}{\log(2)} \end{pmatrix}. \tag{A.40}$$

To derive the covariance matrix we apply multivariate error propagation. In the general case of m functions y_k in n variables $x_1, ..., x_n = \mathbf{x}$, each function y_k is linearized around the mean $\boldsymbol{\mu_x}$

$$y_k \sim y_k(\boldsymbol{\mu_x}) + \sum_{i=1}^{n}(x_i-\mu_i)\left.\frac{\partial y_k}{\partial x_i}\right|_{\boldsymbol{\mu_x}}. \tag{A.41}$$

To derive the covariance matrix **cov** of the functions y_k, the definition of variance is used:

$$\begin{aligned}
\mathbf{cov}_{kl}(\mathbf{y}) &= \langle [y_k(\mathbf{x}) - \langle y_k(\mathbf{x})\rangle][y_l(\mathbf{x}) - \langle y_l(\mathbf{x})\rangle]\rangle \\
&= \langle [y_k(\mathbf{x}) - y_k(\boldsymbol{\mu_x})][y_l(\mathbf{x}) - y_l(\boldsymbol{\mu_x})]\rangle \quad \text{EquationA.41 applied} \\
&= \sum_{k=1}^{m}\sum_{l=1}^{m} \left.\frac{\partial y_k}{\partial x_i}\right|_{\boldsymbol{\mu_x}} \left.\frac{\partial y_l}{\partial x_j}\right|_{\boldsymbol{\mu_x}} \langle (x_i - \mu_i)(x_j - \mu_j)\rangle \\
&= \sum_{k=1}^{m}\sum_{l=1}^{m} \left.\frac{\partial y_k}{\partial x_i}\right|_{\boldsymbol{\mu_x}} \left.\frac{\partial y_l}{\partial x_j}\right|_{\boldsymbol{\mu_x}} \mathbf{cov}_{ij}(\mathbf{x}).
\end{aligned}$$

In a compact notation:

$$\mathbf{cov}(\mathbf{y}) = \mathbf{J}\big|_{\boldsymbol{\mu_x}}\ \mathbf{cov}(\mathbf{x})\ \ \mathbf{J}^T\Big|_{\boldsymbol{\mu_x}}, \tag{A.42}$$

where $\mathbf{cov}(\mathbf{y})$ is the covariance matrix of $\mathbf{y}$, $\mathbf{cov}(\mathbf{x})$ is the covariance matrix of the variables $\mathbf{x}$, and $\mathbf{J}|_{\boldsymbol{\mu_x}}$ is the $m \times n$ Jacobian $J_{ij} = \frac{\partial y_i}{\partial x_i}$ evaluated at the mean of $\mathbf{x}$. Using this method, the covariance matrix of **pd** can be derived:

$$\begin{aligned}
\mathbf{cov_{pd}} &= \begin{pmatrix} \frac{\partial x_{\mathrm{pd}}}{\partial a} & \frac{\partial x_{\mathrm{pd}}}{\partial b} & \frac{\partial x_{\Delta pd}}{\partial \tau} \\ \frac{\partial y_{\mathrm{pd}}}{\partial a} & \frac{\partial y_{\mathrm{pd}}}{\partial b} & \frac{\partial y_{\mathrm{pd}}}{\partial \tau} \end{pmatrix}\Bigg|_{\mu_a,\mu_b,\mu_\tau} \begin{pmatrix} \sigma_a^2 & \sigma_{a,b} & 0 \\ \sigma_{a,b} & \sigma_b^2 & 0 \\ 0 & 0 & \sigma_\tau^2 \end{pmatrix} \begin{pmatrix} \frac{\partial x_{\mathrm{pd}}}{\partial a} & \frac{\partial y_{\mathrm{pd}}}{\partial a} \\ \frac{\partial x_{\mathrm{pd}}}{\partial b} & \frac{\partial y_{\mathrm{pd}}}{\partial b} \\ \frac{\partial x_{\mathrm{pd}}}{\partial \tau} & \frac{\partial y_{\mathrm{pd}}}{\partial \tau} \end{pmatrix}\Bigg|_{\mu_a,\mu_b,\mu_\tau} \\
&= \begin{pmatrix} \sigma^2_{x_{\mathrm{pd}}} & \sigma_{x_{\mathrm{pd}},y_{\mathrm{pd}}} \\ \sigma_{x_{\mathrm{pd}},y_{\mathrm{pd}}} & \sigma^2_{y_{\mathrm{pd}}} \end{pmatrix}
\end{aligned} \tag{A.43}$$

with

$$\sigma^2_{x_{\mathrm{pd}}} = \frac{\log(2)^2\sigma_a^2 + \omega\left(\omega\mu_\tau^2\sigma_b^2 + \omega\mu_b^2\sigma_\tau^2 + \sigma_{a,b}\log(4)\mu_\tau\right)}{\log(2)^2}, \tag{A.44}$$

$$\sigma^2_{y_{\mathrm{pd}}} = \frac{\log(2)^2\sigma_b^2 + \omega\left(\omega\mu_\tau^2\sigma_a^2 + \omega\mu_a^2\sigma_\tau^2 - \sigma_{a,b}\log(4)\mu_\tau\right)}{\log(2)^2} \quad \text{and} \tag{A.45}$$

$$\sigma_{x_{\mathrm{pd}},y_{\mathrm{pd}}} = \frac{\sigma_{a,b}\log(2)^2 - \omega\left(\sigma_{a,b}\omega\mu_\tau^2 + \log(2)\left(\sigma_a^2 - \sigma_b^2\right)\mu_\tau + \omega\mu_a\mu_b\sigma_\tau^2\right)}{\log(2)^2}. \tag{A.46}$$

If the time series is evenly sampled and over an integer number of periods, the covariance matrix has a simpler form:

$$\begin{aligned}
\mathbf{cov_{pd}} &= \begin{pmatrix} \frac{\partial x_{\mathrm{pd}}}{\partial a} & \frac{\partial x_{\mathrm{pd}}}{\partial b} & \frac{\partial x_{\mathrm{pd}}}{\partial \tau} \\ \frac{\partial y_{\mathrm{pd}}}{\partial a} & \frac{\partial y_{\mathrm{pd}}}{\partial b} & \frac{\partial y_{\mathrm{pd}}}{\partial \tau} \end{pmatrix}\Bigg|_{\mu_a,\mu_b,\mu_\tau} \begin{pmatrix} \sigma_{ab}^2 & 0 & 0 \\ 0 & \sigma_{ab}^2 & 0 \\ 0 & 0 & \sigma_\tau^2 \end{pmatrix} \begin{pmatrix} \frac{\partial x_{\Delta pd}}{\partial a} & \frac{\partial y_{\Delta pd}}{\partial a} \\ \frac{\partial x_{\Delta pd}}{\partial b} & \frac{\partial y_{\Delta pd}}{\partial b} \\ \frac{\partial x_{\Delta pd}}{\partial \tau} & \frac{\partial y_{\Delta pd}}{\partial \tau} \end{pmatrix}\Bigg|_{\mu_a,\mu_b,\mu_\tau} \\
&= \begin{pmatrix} \sigma_{ab}^2 + \frac{\omega^2(\mu_b^2\sigma_\tau^2 + \mu_\tau^2\sigma_{ab}^2)}{\log(2)^2} & -\frac{\mu_a\mu_b\sigma_\tau^2\omega^2}{\log(2)^2} \\ -\frac{\mu_a\mu_b\sigma_\tau^2\omega^2}{\log(2)^2} & \sigma_{ab}^2 + \frac{\omega^2(\mu_a^2\sigma_\tau^2 + \mu_\tau^2\sigma_{ab}^2)}{\log(2)^2} \end{pmatrix}.
\end{aligned} \tag{A.47}$$

A.2.2 Sensitivity

In Section 2.5 we describe how a small change of the relative amplitudes A_{prod} and A_{deg} can have a large effect on the phase ϕ_x. This can be quantified as the sensitivity of the phase ϕ_x with respect to changes in the input parameters A_{prod}, A_{deg}, ϕ_{prod} and ϕ_{deg}. **In the following we derive the sensitivity coefficient for the phase ϕ_x as a measure of the sensitivity.** First, we introduce the ratio between the amplitudes of the production term and the degradation rate: $c = A_{\text{prod}}/A_{\text{deg}}$. Then, the phase ϕ_x is calculated from Equation A.17:

$$\phi_x = \arctan 2 \begin{pmatrix} c\cos\left(\phi_{\text{prod}} + \arctan\left(\frac{\omega}{\gamma}\right)\right) - \cos\left(\phi_{\text{deg}} + \arctan\left(\frac{\omega}{\gamma}\right)\right), \\ c\sin\left(\phi_{\text{prod}} + \arctan\left(\frac{\omega}{\gamma}\right)\right) - \sin\left(\phi_{\text{deg}} + \arctan\left(\frac{\omega}{\gamma}\right)\right) \end{pmatrix} . \tag{A.48}$$

We now define the sensitivity coefficient $S_c^{\phi_x}$ as the derivative of the phase ϕ_x with respect to the relation c:

$$S_c^{\phi_x} \overset{\text{def}}{=} \frac{\partial \phi_x}{\partial c} = \frac{\sin(\phi_{\text{deg}} - \phi_{\text{prod}})}{1 + c^2 - 2c\cos(\phi_{\text{deg}} - \phi_{\text{prod}}))} . \tag{A.49}$$

In Figure 2.2C we plotted the sensitivity coefficient for different values of c. Here, a large value means that a small change in the input parameters results in a large change in the phase ϕ_x. The sensitivity is large when production and degradation peak roughly at the same time, i.e. $\phi_{\text{prod}} - \phi_{\text{deg}} \approx 0$. This effect becomes very pronounced when the amplitudes A_{prod} and A_{deg} are similar, i.e. $c \approx 1$. The large sensitivity coefficient implies a phase inversion in this range of c. This means that a small change in the relation of A_{prod} and A_{deg} can cause a phase shift of up to 12 h.

Appendix B: PDE Model - Aging Molecules

In this section I derive the PDE model used in Section 3.3, discuss different rate functions which describe different scenarios and present an analytical solution of the PDE for a specific case. Furthermore, I formulate the inverse problem which can be used to recalculate unmeasured deadenylation rates TAIL-seq data.

B.1 Derivation of the PDE Model

Let $x(t, d')$ be the number of molecules of damage $(d, d + \mathrm{d}d)$ at time t. Therefore $x(t, d)$ is a density function in the damage d. If we integrate $x(t, d)$ over all possible damage values we receive the total molecules concentration at time t:

$$X(t) = \int_{d=0}^{\infty} x(t, d)\ \mathrm{d}d. \tag{B.1}$$

The rate of change in molecules of a given damage interval Δd is defined by

$$\frac{\partial x(t, d)}{\partial t}\Delta d = \begin{bmatrix} + \text{ rate of entry at } d \\ - \text{ rate of departure at } d + \Delta d \\ - \text{ degradation} \end{bmatrix} \tag{B.2}$$

or

$$\frac{\partial x(t, d)}{\partial t}\Delta d = J(t, d) - J(t, d + \Delta d) - v(t, d)x(t, d)\Delta d, \tag{B.3}$$

where $v(t, d)$ is the degradation rate of molecules with damage d at time t. $J(t, d)$ is a flux of molecules with damage d at time t in more detail examined below. Division by Δd yields

$$\frac{\partial x(t, d)}{\partial t} = -\frac{J(t, d) - J(t, d + \Delta d)}{\Delta d} - v(t, d)x(t, d). \tag{B.4}$$

Taking the limit $\Delta d \to 0$ we receive a 'transport equation with advective flux' for the molecules

$$\frac{\partial x(t, d)}{\partial t} = -\frac{\partial J(t, d)}{\partial d} - v(t, d)x(t, d). \tag{B.5}$$

The flux J is not a flux in space as in the original transport equation but a flux in the damage domain and describes the movement of molecules towards higher damage. As a flux in space the damage-dependent flux depends on the molecule density $x(t, d)$ and the speed of damage accumulation $q(t)$. The speed $q(t)$ is defined by the change of damage per time unit $q(t, d) = \frac{\mathrm{d}d}{\mathrm{d}t}$

and the flux J therefore is given by:

$$J(t,d) = x(t,d)q(t,d). \tag{B.6}$$

The model equation Equation B.5 calculates thus to

$$\frac{\partial x(t,d)}{\partial t} = -q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d} - v(t,d)x(t,d). \tag{B.7}$$

This equation is defined for positive d, we assume there is no negative damage, and the damage is steadily increasing, hence we need a boundary condition $x(t,0)$ where molecules with no damage enter the system.

$x(t,d)$ describes the molecule density in damage. For any acceptable solution $x(t,d)$ the integral Equation B.1 is well defined and hence converges which requires that

$$\lim_{d\to\infty} x(t,d) = 0. \tag{B.8}$$

The integral of the model equation EquationB.7 over damage d calculates then

$$\begin{aligned}\int_0^\infty \frac{\partial x(t,d)}{\partial t}\mathrm{d}d &= \int_0^\infty \left(-q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d}\right)\mathrm{d}d\ ,\\ \frac{\mathrm{d}X(t)}{\mathrm{d}t} &= \int_0^\infty \left(-q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d}\right)\mathrm{d}d - \int_0^\infty v(t,d)x(t,d)\mathrm{d}d\ .\end{aligned} \tag{B.9}$$

The first integral on the right hand-side can be reduced by using partial integration

$$\int \left(-q(t,d)\ \frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d}\right)\mathrm{d}d \tag{B.10}$$

$$= -\left(x(t,d)q(t,d) - \int x(t,d)\frac{\partial q(t,d)}{\partial d}\mathrm{d}d\right) - \int x(t,d)\frac{\partial q(t,d)}{\partial d}\mathrm{d}d \tag{B.11}$$

$$= -x(t,d)q(t,d)\ . \tag{B.12}$$

Taking into account borders and Equation B.8 we receive

$$\begin{aligned}\int_0^\infty \left(-q(t,d)\ \frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d}\right)\mathrm{d}d\\ = -\Big[x(t,d)q(t,d)\Big]_0^\infty\\ = x(t,0)q(t,0).\end{aligned} \tag{B.13}$$

The time development of molecule concentration reads therefore (EquationsB.9,B.13)

$$\frac{\mathrm{d}X}{\mathrm{d}t} = x(t,0)q(t,0) - \int_0^\infty v(t,d)x(t,d)\mathrm{d}d\ . \tag{B.14}$$

The equation for a general time development of a molecule concentration X where molecules

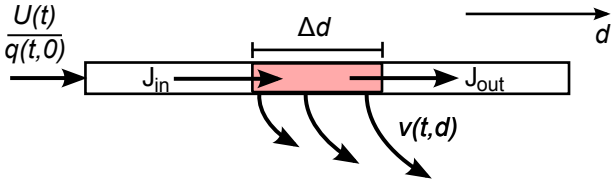


Figure B.1: **Illustration of the model Equation B.17.** Illustration of the model as transport. The molecule density $x(t,d)$ of a given damage interval Δd is determined by the entry J_{in}, the departure J_{out} and degradation $v(t,d)$. Newly synthesized molecules enter the system with the rate $x(t,0) = \frac{U(t)}{q(t,0)}$.

are produced with a production rate k and is degraded with a degradation rate V is given by:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = U - VX. \tag{B.15}$$

We assume that the production rate is time dependent $U = U(t)$. Comparison of EquationB.14 and B.15 gives us then an expression for the boundary condition:

$$x(t,0) = \frac{U(t)}{q(t,0)}. \tag{B.16}$$

With that we received the model Equation 3.6 which describes a molecule density in damage where molecules are born with zero damage, accumulate damage over time described by a flux in d-direction and get eventually degraded (see Fig.B.1):

$$\begin{aligned} \frac{\partial x(t,d)}{\partial t} &= -q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d} - v(t,d)x(t,d) \\ x(t,0) &= \frac{k(t)}{q(t,0)} \end{aligned} \tag{B.17}$$

To calculate from the molecule density $x(t,d)$ the actual molecule concentration of molecules within a certain damage range $[d, d+\Delta d]$ one has to take the integral

$$X_{d,d+\Delta d}(t) = \int_d^{d+\Delta d} x(t,d)\mathrm{d}d\,. \tag{B.18}$$

B.2 Rates - Personalize your Model

In the following we take a closer look at the three rates we use in our model: $k(t)$ is the production rate, $v(t,d)$ is connected to the degradation of molecules, and $q(t,d)$ describes the speed of damage accumulation.

With a specific mathematical formulation of the rates the very general model Equation B.17

can describe specific biological situations. Additionally, besides the mathematical description of biological processes, we need to make a connection to measurable quantities. In this section I will draw a connection from the damage-dependent degradation rate $v(t, d)$ to the general molecule degradation rate $V(t)$ which is damage independent and usually measured in experiments when investigating molecule half lives. Furthermore, I will present some general ideas how all rates, $k(t)$, $q(t, d)$ and $v(t, d)$ can be formulated in the context of circadian biology and in the examples presented in this thesis where I used the model to describe such diverse phenomena as protein damage and poly(A) tail deadenylation.

The production rate describes only a production of undamaged proteins ($d = 0$). Hence, in contrast to the degradation rate the production rate $k(t)$ of the molecule concentration $X(t)$ therefore already only depends on time

$$U(t) = k(t) \tag{B.19}$$

and is the production rate which would be measured. For example if describing deadenylation of poly(A) tails the production rate is constituted by transcriptional activity.

To get a meaningful degradation rate, i.e. an experimentally measurable rate, we compare Equation B.14 and Equation B.15 and extract the term which describes the degradation of molecules:

$$V(t)X(t) = \int_0^\infty v(t, d)x(t, d)\mathrm{d}d. \tag{B.20}$$

$V(t)$ is the damage-independent degradation rate of molecules and connected to the half-life τ via $\tau(t, d) = \ln(2)/V(t)$.

How should all these rates $k(t)$, $q(t, d)$ and $v(t, d)$ be formulated? Describing phenomena with a circadian rhythm demands a time dependence in each rate. I will model this time dependence as periodic cosine-function with mean M, relative amplitude A_{rel} and a phase shift ϕ

$$f(t) = M\left(1 + A_{\mathrm{rel}}\cos(\omega t - \phi)\right). \tag{B.21}$$

Since production depends only on time it is described only by this cosine function with specific mean k, relative amplitude A_{prod} and phase ϕ_{prod}:

$$k(t) = k\left(1 + A_{\mathrm{prod}}\cos(\omega t - \phi_{\mathrm{prod}})\right) \tag{B.22}$$

The rate of damage accumulation is also circadian but additionally it can depend on the damage d:

$$q(t, d) = c\left(1 + A_{\mathrm{dam}}\cos(\omega t - \phi_{\mathrm{dam}})\right) \cdot f(d). \tag{B.23}$$

The function $f(d)$ modulates the mean damage accumulation c and should hence do not exceed a certain value range, e.g. $f(d) \in [0, 1]$ to be reasonable (within the measured range) damage d.

To reproduce the data from Chang et al. [252] I used a simple exponential function for the damage dependency:

$$q(d) = e^{-0.1d}. \tag{B.24}$$

This describes that deadenylation slows down exponentially with shortened poly(A) tail. How-

ever, any other description could be plausible, for example a step-like function describing two different (constant) deadenylation speeds of two different deadenylases acting only at specific poly(A) tail lengths.

Also the degradation rate can be damage dependent. To describe RNA molecules which are degraded only with a short poly(A) tail I used a tunable step-like function $v(t, d)$:

$$v(t,d) = \gamma \cdot \underbrace{(1 + A_{\text{deg}} \cos(\omega t - \phi_{\text{deg}}))}_{\text{time-dependence}} \cdot \underbrace{\frac{\left(\frac{d-\beta}{s}\right)^h}{1 + \frac{d-\beta}{s}} \cdot H(d-\beta)}_{\text{damage-dependence}}. \tag{B.25}$$

Here, the function is defined by the mean degradation rate γ and two terms, the first only dependent on time, the second only dependent on damage. While the time-dependence has the same structure as time-dependence defined in Equation B.21, the part with damage-dependence deserves more explanation. We assume that degradation eventually get switched on at a certain threshold of accumulated damage or shortened poly(A) tail. The damage-dependence part has to reflect this behavior. It consists of a sigmoid Hill-function to reflect a switching behavior. This function is defined by the parameter h and s, the first is the so-called Hill factor h and defines the steepness of the slope, the second scales the slope to a given range. The Hill-function is further shifted in d-direction by the parameter β and the Heaviside-function $H(d-\beta)$ assures that the shifted Hill-function takes only values higher than zero. In Fig. B.2 the damage-dependence of the rate $v(t, d)$ is plotted for different parameter values.

How can we translate the proposed rate $v(t, d)$ to the measurable degradation rate $V(t)$ defined by Equation B.15. Inserting $v(t, d)$ into Equation B.20 gives:

$$V(t,d)X(t) = \int_0^\infty \gamma \left(1 + A_{\text{deg}} \cos(\omega t - \phi_{\text{deg}})\right) \frac{\left(\frac{d-\beta}{s}\right)^h}{1 + \frac{d-\beta}{s}} H(d-\beta) x(t,d) \mathrm{d}d \tag{B.26}$$

$$V(t,d)X(t) = \gamma \left(1 + A_{\text{deg}} \cos(\omega t - \phi_{\text{deg}})\right) \int_0^\infty \frac{\left(\frac{d-\beta}{s}\right)^h}{1 + \frac{d-\beta}{s}} H(d-\beta) x(t,d) \mathrm{d}d. \tag{B.27}$$

The integral in Equation B.27 is now defined by the damage-dependence of $v(t, d)$ and the molecule density $x(t, d)$. As the integral is not solvable in the present form we examine the limit case for a small scaling factor s and a high Hill factor h. This limit case will result in a step-like function and the damage dependence of $v(t, d)$ reduces to

$$v(d) = \frac{\left(\frac{d-\beta}{s}\right)^h}{1 + \frac{d-\beta}{s}} H(d-\beta) \quad \xrightarrow[s \ll 1,\, h \gg 1]{} \quad \begin{cases} 0 & \text{for } d < \beta \\ 1 & \text{for } d \leq \beta \end{cases} \tag{B.28}$$

Inserting this extreme case EquationB.28 in EquationB.27 gives an estimate for the degradation

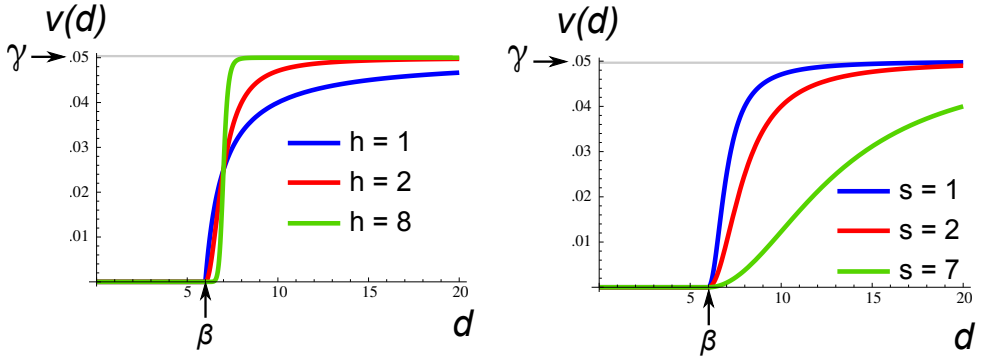


Figure B.2: **The damage dependence of $v(t,d)$ (Equation B.25) for different parameter values.** The Hill factor h defines the steepness of the step, parameter s scales the slope to a range. Parameter, if not indicated: $s = 1$, $h = 2$, $\beta = 6$, $\gamma = 0.05$.

term which the molecule concentration $X(t)$ experiences:

$$V(t)X(t) = \gamma\left(1 + A_{\text{deg}}\cos(\omega t - \phi_{\text{deg}})\right)\int_{\beta}^{\infty} x(t,d)\mathrm{d}d \tag{B.29}$$

B.3 Analytical Solution of the PDE

Here, I will demonstrate the derivation of an analytical solution for the PDE. However, a closed solution is only possible for damage-independent rates. In this section I denote

$$\frac{\partial x(t,d)}{\partial t} = x_t \quad \text{and} \tag{B.30}$$

$$\frac{\partial x(t,d)}{\partial d} = x_d. \tag{B.31}$$

The PDE with time-dependent production and degradation rates and constant damage accumulation rate reads then:

$$0 = x_t + cx_d + v(t)x \quad \text{with}$$

$$x(t,0) = \frac{k}{c}\left(1 + A_{\text{prod}}\cos(\omega t - \phi_{\text{prod}})\right) quad\text{and} \tag{B.32}$$

$$v(t) = \gamma\left(1 + A_{\text{deg}}\cos(\omega t - \phi_{\text{deg}})\right)\left(e^{\beta d} - 1\right).$$

To solve this equation I use the method of characteristics, where one variable is limited by basically changing the coordinate system such that the eliminated variable constitutes the new coordinate system. It is only applicable for first order linear PDEs with constant coefficients. In this case it means, in order to use the method the damage accumulation rate is required to be constant. I learned this standard technique for solving PDEs from the well written script on

PDEs by Victor Grigoryan (`http://web.math.ucsb.edu/~grigoryan/124A/lectures.html`). The new variables ξ and η then read

$$\xi = t + cd \tag{B.33}$$
$$\eta = ct + d. \tag{B.34}$$

The old variables expressed by the new ones:

$$t = \frac{\xi + c\eta}{1 + c^2} \tag{B.35}$$
$$d = \frac{c\xi - \eta}{1 + c^2}. \tag{B.36}$$

With this new variables the PDE turns into an ODE:

$$0 = (1 + c^2)x_\xi + v(\xi, \eta)x \tag{B.37}$$
$$v(\xi, \eta) = \gamma \left(1 + A_{\text{deg}} \cos\left(\omega\xi + c\eta 1 + c^2 - \phi_{\text{deg}}\right)\right) \left(e^{\beta \frac{c\xi - \eta}{1+c^2}} - 1\right) \tag{B.38}$$

This homogeneous ODE can be solved by separating the variables. After some integration and retransformation to the old variables one obtains for the general solution of the homogeneous ODE

$$x(t, d) = f(ct - d) \exp\left\{\gamma\left[- \frac{1}{1 + c^2}(t + cd) - A_{\text{deg}}\left[\sin\left(\omega t - \phi_{\text{deg}}\right)\left(-\frac{1}{\omega} + \frac{\omega}{(\beta c)^2 + \omega^2} e^{\beta d}\right) + \right.\right.\right.$$
$$\left.\left.\left. + + \cos\left(\omega t - \phi_{\text{deg}}\right)\left(\frac{\beta c}{(\beta c)^2 + \omega^2} e^{\beta d}\right)\right]\right]\right\} \tag{B.39}$$

The unknown function $f(ct - d)$ is determined by the initial condition:

$$x(t, 0) = \frac{k}{c}\left(1 + A_{\text{prod}} \cos(\omega t - \phi_{\text{prod}})\right) = k(ct) = \tag{B.40}$$
$$= f(ct) \exp\left\{\gamma\left[- \frac{1}{1 + c^2} t - A_{\text{deg}}\left[\sin\left(\omega t - \phi_{\text{deg}}\right)\left(-\frac{1}{\omega} + \frac{\omega}{(\beta c)^2 + \omega^2}\right) + \right.\right.\right.$$
$$\left.\left.\left. + \cos\left(\omega t - \phi_{\text{deg}}\right)\left(\frac{\beta c}{(\beta c)^2 + \omega^2}\right)\right]\right]\right\} \tag{B.41}$$
$$= f(ct) g(ct). \tag{B.42}$$

It follows

$$f(ct) = \frac{k(ct)}{g(ct)} \quad \text{and} \tag{B.43}$$

$$f(ct-d) = \frac{k(ct-d)}{g(ct-d)}. \tag{B.44}$$

With this one receives the final solution for the PDE Equation B.32:

$$\begin{aligned} x(t,d) =& \frac{k}{c}\left(1 + A - \operatorname{prod}\cos\left(\omega t - \omega\frac{d}{c} - \phi_{\text{prod}}\right)\right)\cdot \\ & \cdot \exp\left\{ -\gamma\left[\frac{1}{\beta c}\left(e^{-\beta d}-1\right) - \frac{d}{c} + \right.\right. \\ & + A_{\text{deg}}\left[\sin\left(\omega t - \phi_{\text{deg}}\right)\left(-\frac{1}{\omega} + \frac{\omega}{(\beta c)^2+\omega^2}e^{\beta d} + \frac{(\beta c)^2}{\omega\left((\beta c)^2+\omega^2\right)}\cos(\omega d/c) - \frac{\beta c}{(\beta c)^2+\omega^2}\sin(\omega d/c)\right)\right. \\ & \left.\left.\left. + \cos\left(\omega t - \phi_{\text{deg}}\right)\left(\frac{\omega}{(\beta c)^2+\omega^2}e^{\beta d} - \frac{(\beta c)^2}{\omega\left((\beta c)^2+\omega^2\right)}\cos(\omega d/c) + \frac{\beta c}{(\beta c)^2+\omega^2}\sin(\omega d/c)\right)\right]\right]\right\} \end{aligned} \tag{B.45}$$

B.4 Recalculation of Deadenylation Rates: Inverse Problem

I use the model Equation B.17 in Section 3.3.3 to describe the deadenylation of poly(A) tails, stabilizing elements at the 3'end of an mRNA. A new sequencing techniques allows genome-wide measurement of poly(A) tails. With this data, in the model description constituting $x(t,d)$, it is possible to recalculate unmeasured mRNA specific deadenylation rates $q(t,d)$ by formulating an inverse problem as I will outline in the following.

The model describes development of the concentration of mRNA $x(t,d)$ with certain poly(A) tail lengths d in time t. In this description $d=0$ means an intact, unshortened poly(A) tail of 250 adenosides, $x(t,d=250)$ describes the concentration of mRNA with completely removed poly(A) tail. These mRNA with their poly(A) tails are transcribed with rate $k(t)$ and degraded with $v(t,d)$. The poly(A) tail is shortened with rate $q(t,d)$. These processes are captured by the PDE model from the previous section:

$$\begin{aligned} \frac{\partial x(t,d)}{\partial t} &= -q(t,d)\frac{\partial x(t,d)}{\partial d} - x(t,d)\frac{\partial q(t,d)}{\partial d} - v(t,d)x(t,d) \\ x(t,0) &= \frac{k(t)}{q(t,0)} \end{aligned} \tag{B.46}$$

The TAIL-seq data gives poly(A) tail length distributions for each gene, in the model description a time-independent solution of the model $x(d)$. To use these data to recalculate poly(A) tail length-dependent deadenylation rates, in the model description $q(d)$, we need to discuss the influence and make assumptions with regards to all other parameters. It is well established that a poly(A) tail of about 250 nucleotides is added to each mRNA right after completion of its transcription [247] and mRNA are mostly only degraded after complete removal of their poly(A) tails [187]. For the model this means mRNAs are born only with a complete poly(A) tail,

consequently the transcription rate enters the model only as boundary condition. Furthermore, mRNAs are not degraded if the poly(A) tail is longer than 20 nucleotides and hence for $d < 230$ the degradation rate equals zero, $v(t, d < 230) = 0$. Furthermore, to formulate an inverse problem with an unique solution I need to assume that poly(A) tails are not relengthened. This is true for most RNA, however it is known that cytosolic relengthening is possible. For example, some mRNA of housekeeping genes are stored in stress granules, where mRNA are translationally silenced via a shortened poly(A) tail and upon stress the translation is reactivated through poly(A) tail relengthening [259]. Since one TAIL-seq measurement captures one snapshot in time the model description does not include time dependence. To capture a poly(A) tail length distribution which changes in time, e.g. a poly(A) tail length distribution which is influenced by the circadian deadenylase Nocturnin [166], one would need a time series of these measurements and apply the following procedure to each TAIL-seq data separately.

Applying all the above mentioned reduces the Equation B.46 for $d < 230$ to

$$\begin{aligned} 0 &= -q(d)\frac{\mathrm{d}x(d)}{\mathrm{d}d} - x(d)\frac{\mathrm{d}q(d)}{\mathrm{d}d} \\ x(0) &= \frac{k}{q(0)}. \end{aligned} \tag{B.47}$$

This model describes the development of a poly(A) tail length of one mRNA species for $d < 230$. The unknown function $x(d)$ solves this equation. However, with the TAIL-seq data $x(d)$ is known and the deadenylation rate is the unknown function. To recalculate the length-dependent deadenylation rate from the TAIL-seq data I discretize the derivatives, for example:

$$\frac{\mathrm{d}q(d)}{\mathrm{d}d} = \frac{q_k - q_{k-1}}{\Delta d}, \tag{B.48}$$

where the step size Δd equals one adenoside, the smallest reasonable solution in this description. With this I gain from Equation B.47 a system of linear equations

$$x_0 = \frac{k}{q_0} \tag{B.49}$$

$$0 = -q_k\frac{x_k - x_{k-1}}{\Delta d} + x_k\frac{q_k - q_{k-1}}{\Delta d}\ , k = 0...220. \tag{B.50}$$

This system of linear equation with given x_k from the TAIL-seq data is solvable for q_k except for a proportionality factor k, discrete values of the deadenylation rate $q(d)$. To relate q_k to absolute values one would need additionally information on the transcriptional activity k.

In the main text, Section 3.3.3, the discretization and recalculation of $q(d)$ has been done on a simulated model which resembled measured data. To apply it on real data would follow the same procedure. However, I expect that the data has to be smoothed beforehand.

Appendix C: Data Processing

C.1 Sequencing Data

C.1.1 Read Quantification

For the liver data set, already published in Menet *et al.* [5], raw data was downloaded from the public domain at GEO with the accession number GSE36916 (`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36916`). The raw data for mouse kidney was provided by Roman-Ulrich Müller from University of Cologne.

All raw data was annotated using Bowtie2 [260].

Read quantification was done using an R-script which is geared to the vignette "Overlap encodings" of the package "GenomicAlignments" [261]. Read counts were normalized to Reads Per Kilobase per Million mapped reads (RPKM).

In short, for each gene a list of the genomic ranges of exons and introns were extracted from the reference genome `TxDb.Mmusculus.UCSC.mm9.refGene`. For liver, two sequencing data sets were available, nascent-sequencing and common RNA sequencing. Reads from nascent-seq which overlapped with exons of a gene were counted as its transcriptional activity, reads from RNA sequencing which overlapped with exons of a gene were counted as RNA abundance. For kidney, information on both transcriptional activity and mRNA abundance came from one RNA -sequencing data set. Aligned reads from introns of a gene were counted as its transcriptional activity. Aligned reads from exons were counted as RNA abundance.

The threshold for expressed genes were set after inspection of histograms of read counts according to the following table.

data set	Expressed genes are genes with mean read count larger than
kidney, transcriptional activity	2^{-7}.
kidney, RNA abundance	2^{-5}.
liver, transcriptional activity	2^{-5}.
liver, RNA abundance	2^{-3}.

From now on I only consider expressed genes without an explicit statement.

C.1.2 Circadian Genes

To test if time series of either transcriptional activity or RNA abundance are rhythmic I tested each time series with RAIN [200] by using R-package "rain" (Version 1.4.0) and used Benjamini-Hochberg correction to calculate for each time series a False Discovery Rate (FDR). Additionally, I fitted a cosine curve to each time series using the R-package "HarmonicRegression" [179].

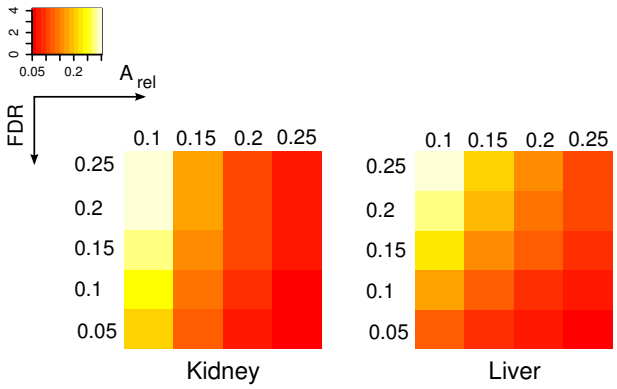


Figure C.1: **Proportion of genes which are classified as circadian.** RNA abundances were classified as circadian or not circadian with different threshold values. Color-coded are percentages of circadian RNA from all expressed RNA. FDR comes from RAIN, relative amplitude from cosine fit.

From this fit I obtained the relative amplitudes. In Fig. C.1 the proportion of genes which are classified as circadian of all expressed RNA abundances for different thresholds are shown. It is interesting to see that mainly the relative amplitude is the determining factor if gene is classified as circadian. If not stated elsewhere I classified time series with a FDR$\leq$ 0.25 and a relative amplitude > 0.1 as circadian.

C.1.3 Estimate Uncertainty of Cosine Fit

A rhythmic abundance of a species x is measured by a time series consisting of n points. The time series is fitted with a linear model

$$x_t = A\cos(\omega t) + B\sin(\omega t) + a_t \tag{C.1}$$

where $\{a_t\}$ are random and uncorrelated errors with a standard deviation σ:

$$\begin{aligned} E(a_t) &= 0 \\ E(a_t a_t') &= \begin{cases} \sigma^2, & t = t' \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{C.2}$$

C.1.4 Covariance Matrix of Fitting Parameters

In matrix form the fitting problem Equation (C.1) can be stated as

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X} \tag{C.3}$$

with

$$\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \cos(\omega t_1) & \sin(\omega t_1) \\ \cos(\omega t_2) & \sin(\omega t_2) \\ \vdots & \\ \cos(\omega t_n) & \sin(\omega t_n) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} A \\ B \end{pmatrix} \tag{C.4}$$

In the method of linear least squares we are now seeking for parameters A and B which minimize the sum of squared vertical distances $S(A, B)$ between the observed responses in the dataset and the responses predicted by the linear approximation:

$$S(A, B) = \sum_{i=1}^{n} (x_i - A\cos(\omega t) - B\sin(\omega t))^2 = (\mathbf{y} - \boldsymbol{\beta}\mathbf{X})^T(\mathbf{y} - \boldsymbol{\beta}\mathbf{X}) \tag{C.5}$$

This function has a global minimum at $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$. The covariance matrix of the fitting parameters $\mathbf{cov}_{\boldsymbol{\beta}}$ is given by

$$\mathbf{cov}_{\boldsymbol{\beta}} = \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{C.6}$$

where σ^2 is the standard deviation given in Equation (C.2). The time series are measured at equally distributed along the period time points and can consists either of fully measured periods or fully measured periods plus additional time points. Assuming, we have m fully measured periods with N time points each and k additional time points. An example would be a circadian time series (period = 24) measured at time points 0, 5, 11, 17, 23, 29. This would mean 6 time points in total, $m = 1$ fully measured period of $N = 4$ time points per period and $k = 1$ additional time point.

We can now rewrite the matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{pmatrix} \cos\left(\frac{2\pi}{N}0\right) & \sin\left(\frac{2\pi}{N}0\right) \\ \cos\left(\frac{2\pi}{N}1\right) & \sin\left(\frac{2\pi}{N}1\right) \\ \vdots & \\ \cos\left(\frac{2\pi}{N}m \cdot N + k\right) & \sin\left(\frac{2\pi}{N}m \cdot N + k\right) \end{pmatrix} \tag{C.7}$$

To calculate the covariance matrix of the fitting parameters Equation (C.6) we first look at the matrix $\mathbf{X}^T\mathbf{X}$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \cos\left(\frac{2\pi}{N}0\right) & \cos\left(\frac{2\pi}{N}1\right) & \dots & \cos\left(\frac{2\pi}{N}m \cdot N + k\right) \\ \sin\left(\frac{2\pi}{N}0\right) & \sin\left(\frac{2\pi}{N}1\right) & & \sin\left(\frac{2\pi}{N}m \cdot N + k\right) \end{pmatrix} \cdot \begin{pmatrix} \cos\left(\frac{2\pi}{N}0\right) & \sin\left(\frac{2\pi}{N}0\right) \\ \cos\left(\frac{2\pi}{N}1\right) & \sin\left(\frac{2\pi}{N}1\right) \\ \vdots & \\ \cos\left(\frac{2\pi}{N}m \cdot N + k\right) & \sin\left(\frac{2\pi}{N}m \cdot N + k\right) \end{pmatrix}$$

$$= \begin{pmatrix} CC & CS \\ CS & SS \end{pmatrix} \tag{C.8}$$

with following symbols:

$$CC = \sum_{i=0}^{mN+k-1} \cos^2\left(\frac{2\pi}{N}i\right) \tag{C.9}$$

$$SS = \sum_{i=0}^{mN+k-1} \sin^2\left(\frac{2\pi}{N}i\right) \tag{C.10}$$

$$CS = \sum_{i=0}^{mN+k-1} \cos\left(\frac{2\pi}{N}i\right)\sin\left(\frac{2\pi}{N}i\right) \tag{C.11}$$

All these sums can be divided into two parts of fully measured periods and the rest. As an example I show this for the entry CC Equation C.9:

$$\begin{aligned} CC &= \sum_{i=0}^{mN+k-1} \cos^2\left(\frac{2\pi}{N}i\right) \\ &= \sum_{i=0}^{mN-1} \cos\left(\frac{2\pi}{N}i\right) + \sum_{i=mN}^{mN+k-1} \cos^2\left(\frac{2\pi}{N}i\right) \\ &= CC_{\text{full periods}} + CC_{\text{rest}} \end{aligned} \tag{C.12}$$

The sums of fully measured periods can be strongly simplified. The sum of squared cosines reduces to

$$\begin{aligned} \sum_{i=0}^{mN-1} \cos^2\left(\frac{2\pi}{N}i\right) &= \sum_{i=0}^{mN-1} \frac{1+\cos\left(\frac{4\pi}{N}i\right)}{2} \quad \text{used trigonometric identity} \\ &= \frac{mN}{2} + \frac{1}{2}\underbrace{\sum_{i=0}^{mN-1} \cos\left(\frac{4\pi}{N}i\right)}_{\substack{= 0,\ \text{sum of } mN \text{ cosines} \\ \text{which are evenly distributed on a } 4\pi\text{-period}}} \\ &= \frac{mN}{2} = CC_{\text{full periods}}. \end{aligned} \tag{C.13}$$

We use this result and the Pythagorean identity to calculate the sum of squared sines:

$$\begin{aligned} mN &= \sum_{i=0}^{mN-1} 1 \\ &= \sum_{i=0}^{mN-1} \left(\cos^2\left(\frac{2\pi}{N}i\right) + \sin^2\left(\frac{2\pi}{N}i\right)\right) \quad \text{used Pythagorean identity} \\ &= \sum_{i=0}^{mN-1} \cos^2\left(\frac{2\pi}{N}i\right) + \sum_{i=0}^{mN-1} \sin^2\left(\frac{2\pi}{N}i\right) \\ &= \frac{mN}{2} + \sum_{i=0}^{mN-1} \sin^2\left(\frac{2\pi}{N}i\right) \quad \text{used result in Equation C.13} \end{aligned}$$

$$\Leftrightarrow \sum_{i=0}^{mN-1} \sin^2\left(\frac{2\pi}{N}i\right) = \frac{mN}{2} = SS_{\text{full periods}}. \tag{C.14}$$

The sum of the mixed product reduces to

$$\begin{aligned} \sum_{i=0}^{mN-1} \cos\left(\frac{2\pi}{N}i\right)\sin\left(\frac{2\pi}{N}i\right) &= \frac{\sin\left(\frac{4\pi}{N}i\right) + \sin(0)}{2} \quad \text{used trigonometric identity} \\ &= \underbrace{\sum_{i=0}^{mN-1} \frac{\sin\left(\frac{4\pi}{N}i\right)}{2}}_{\substack{= 0, \text{ sum of } mN \text{ sines} \\ \text{which are evenly dis-} \\ \text{tributed on a } 4\pi\text{-period}}} \\ &= 0 = CS_{\text{full periods}} \end{aligned} \tag{C.15}$$

Taken together the results from Equations (C.8-C.15) the matrix $\mathbf{X}^T\mathbf{X}$ transforms to:

$$\begin{aligned} \mathbf{X}^T\mathbf{X} &= \begin{pmatrix} CC & CS \\ CS & SS \end{pmatrix} \\ &= \begin{pmatrix} CC_{\text{full periods}} + CC_{\text{rest}} & CS_{\text{full periods}} + CS_{\text{rest}} \\ CS_{\text{full periods}} + CS_{\text{rest}} & SS_{\text{full periods}} + SS_{\text{rest}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{mN}{2} + CC_{\text{rest}} & CS_{\text{rest}} \\ CS_{\text{rest}} & \frac{mN}{2} + SS_{\text{rest}} \end{pmatrix} \end{aligned} \tag{C.16}$$

The covariance matrix of the parameters $\mathbf{cov}_\beta$ of only full measured periods ($CC_{\text{rest}} = SS_{\text{rest}} =$

$CS_{\text{rest}} = 0$) is then given by Equation (C.6):

$$\begin{aligned}\mathbf{cov}_{\beta,\text{ full}} &= \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \\ &= \sigma^2 \begin{pmatrix} \frac{mN}{2} & 0 \\ 0 & \frac{mN}{2} \end{pmatrix}^{-1} \\ &= \sigma^2 \begin{pmatrix} \frac{2}{mN} & 0 \\ 0 & \frac{2}{mN} \end{pmatrix}\end{aligned} \tag{C.17}$$

and for time series with fully measured periods and some additional time points the covariance matrix $\mathbf{cov}_\beta$ reads

$$\begin{aligned}\mathbf{cov}_\beta &= \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \\ &= \sigma^2 \begin{pmatrix} \frac{mN}{2} + CC_{\text{rest}} & CS_{\text{rest}} \\ CS_{\text{rest}} & \frac{mN}{2} + SS_{\text{rest}} \end{pmatrix}^{-1} \\ &= \frac{\sigma^2}{\left(\frac{mN}{2} + CC_{\text{rest}}\right)\left(\frac{mN}{2} + SS_{\text{rest}}\right) - CS_{\text{rest}}^2} \begin{pmatrix} \left(\frac{mN}{2} + SS_{\text{rest}}\right) & -CS_{\text{rest}} \\ -CS_{\text{rest}} & \left(\frac{mN}{2} + CC_{\text{rest}}\right) \end{pmatrix}\end{aligned} \tag{C.18}$$

with following abbreviations:

$$\begin{aligned} CC_{\text{rest}} &= \sum_{i=mN}^{mN+k-1} \cos^2\left(\frac{2\pi}{N}i\right), \\ SS_{\text{rest}} &= \sum_{i=mN}^{mN+k-1} \sin^2\left(\frac{2\pi}{N}i\right) \text{ and} \\ CS_{\text{rest}} &= \sum_{i=mN}^{mN+k-1} \cos\left(\frac{2\pi}{N}i\right)\sin\left(\frac{2\pi}{N}i\right). \end{aligned} \tag{C.19}$$

C.2 Half-Lives

Half-lives were collected from two studies: Friedel *et al.* [212] and Schwanhäusser *et al.* [213], where mRNA half-lives in NIH3T3 mouse fibroblasts were assayed. In these two studies, a technique based on labeling of newly synthesized mRNA was used, resulting in significantly more accurate half-life estimates than obtained by older technique based on transcriptional arrest induced by actinomycin [262]. These data were merged on MGI gene symbol, a geometric mean (equivalent to mean of the inverse half-lives or, in other words, degradation rates) was used for the cases several transcripts mapped to the same gene. In total, 9595 mRNA half-lives mappable to MGI symbols were assayed by at least one of these two studies, 3308 were assayed in both. A simple linear least squares fit $\log(t_{1/2,\text{ Friedel}}) = a + b\log(t_{1/2,\text{ Schwanhäusser}})$ yielded estimates $a = -0.22$ and $b = 0.98$. Since b is within the error margin of 1 the fit is excellent, and we chose to correct the Schwanhäusser *et al.* [213] data for the offset by multiplication with the

factor 10^a. For the 3308 cases where half-lives were obtained in both studies, unbiased estimates of their variances were made, using the corrected values of Schwanhäusser *et al.* [213].

C.3 UTRs

In Section 2.7.1 I investigated the 5' and 3' UTRs of genes positively tested for rhythmic post-TRXreg in mouse liver and kidney. For this I reprocessed the annotated reads and quantified all reads not on gene annotation but on transcript annotation, see Section C.1.1, followed by finding circadian transcripts using a q-value cut-off obtained by RAIN and an amplitude cut-off (FDR<0.25, $A_{\text{rel}} \leq 0.1$), see Section C.1.2. Then I applied to these transcripts the PA-test to find transcripts under rhythmic post-transcriptional control as I also already did for sequencing data quantified using the gene annotation.

From these transcripts in liver and kidney, which are positively tested for rhythmic post-transcriptional regulation I retrieved the UTRs. A complete list of all 5'UTRs and 3'UTRs of mouse transcripts was obtained using the R-package GenomicRanges [261] together with the reference genome `TxDb.Mmusculus.UCSC.mm9.refGene`. To use these lists it was necessary to map the RefSeq identifier gained during transcript quantification to UCSC transcript identifier. For that I used a list downloaded from the USCS Genome Browser, complete url: `http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/knownToRefSeq.txt.gz`. When inspecting this list I noticed that mapping of RefSeq, already a transcript identifier, to UCSC identifier was a one to many mapping and upon inspection these UCSC mapping to one RefSeq were also annotated with the same UTRs. Consequently, using the unprocessed list of UCSC identifier would result in several annotation of many UTRs for no identifiable reason. I therefore kept for each RefSeq identifier only one UCSC identifier.

I used this list of UCSC identifiers together with the lists of 5' and 3'UTRs to get the UTR lengths. For each organ and each UTR (either 5' or 3') I used two reference lists of transcripts:

- all expressed transcripts
- circadian transcripts positively tested for rhythmic post-transcriptional regulation (FDR<0.25 obtained from PA-test [179]).

C.4 RNA Binding Proteins

In Section 2.7.2 I search for rhythmically expressed RNA binding proteins (RBPs). In order to do that I use a combination of three curated lists of RBPs in mouse:

- `http://rbpdb.ccbr.utoronto.ca`: a collection of experimental observations of RNA-binding sites, both in vitro and in vivo, manually curated from primary literature by Cook *et al.* [263]
- `http://cisbp-rna.ccbr.utoronto.ca`: a Catalog of Inferred Sequence Binding Preferences of RNA binding proteins by Ray *et al.* [264]

- `http://rbpmap.technion.ac.il`: a database of 114 human/mouse RBPs together with their motifs by Paz *et al.* [265]

This gives list of 463 RBPs in mouse. To find circadianly expressed RBPs I identify for each RBP the transcript in liver and kidney and choose these RBPs whose mRNA abundance have a strong circadian rhythm, i.e. FDR$\leq$ 0.25 obtained from RAIN and $A_{\text{rel}} > 0.2$. Furthermore, I exclude all RBPs whose half-lives has been measured by Schwanhäusser *et al.* [213] since all of these have a half-life are much longer than 15 hours and I expect for these RBPs that rhythms in protein level has vanished, see Section 2.1. With that I receive 15 and 25 RBPs with a rhythmic transcript in kidney and liver, respectively. 4 of them are rhythmic in both organs. A complete list of these RBPs can be found in Section F.4.

C.4.1 Enrichment of RBP Binding Sites

In Section 2.7.3 I searched for RNA binding sites in the 5' and 3'UTRs of transcripts. I then test if RBP binding sites of specific RBPs are enriched in genes positively tested for rhythmic postTRXreg.

To find the number of RNA binding sites for each transcript I again used the transcripts quantification described in Section C.3 and downloaded the UTR-sequences from Ensembl BioMart Browser (`www.ensembl.org/biomart/martview`) [266] accessed by ENSEMBLE Transcript identifier. These sequences I used in the online tool "RBPmap (Mapping Binding Sites of RNA Binding Protein)" [265]. This tool provides a curated list of 114 human/mouse RBP binding motifs. I allowed to search for occurrence of any of these motifs in the uploaded sequences together with options "High stringency", which applies a very low p-value cut-off ($p < 0.001$).

I identified in this way RBP binding motifs in two transcript sets for each organ:

- all circadian transcripts (FDR$\leq$ 0.25 obtained from RAIN and $A_{\text{rel}} > 0.1$)
- circadian genes positively tested for rhythmic postTRXreg (FDR$\leq$ 0.25 obtained from PA-test).

For each transcript I annotated if a RBP binding motif of a specific RBP in the UTRs occurred. In this way I received how many transcript UTR contained a binding site for each RBP. These counts of binding site occurrences I used to test for enrichment of transcripts positively tested for rhythmic postTRXreg against all circadian transcripts. The enrichment test was done using Fisher's Exact test implemented in R (function fisher.test).

C.5 Gene Function and Enrichment of Gene Functions

In Section 2.7.3 I also tested for gene function enrichment. For that I extracted different subsets of genes and used them as test set and background set in the DAVID Function Annotation Tool. I accessed gene functions via the genes' Entrez gene ID.

I investigated following subsets for each organ (see also text and figure in Section 2.7.3):

subset	used as
all circadian genes	background1 (bg1)
all genes tested for rhythmic postTRXreg	testset with bg1 / background2 (bg2)
testset1	testset with bg2
genes from testset1 + pdp btw 0 & 12 h	testset with bg2
genes from testset1 + pdp btw 12 & 0 h	testset with bg2
genes from testset1 + pdp btw 6 & 18 h	testset with bg2
genes from testset1 + pdp btw 18 & 6 h	testset with bg2
genes from testset1 + pdp btw 0 & 6 h	testset with bg2
genes from testset1 + pdp btw 6 & 16 h	testset with bg2
genes from testset1 + pdp btw 16 & 24 h	testset with bg2

Abbreviations:
all circadian genes = all expressed genes (either mRNA or transcriptional activity) with RAIN FDR$\leq$ 0.25 and $A_{\text{rel}} > 0.1$
testset1 = all circadian genes positivly tested for rhythmic postTRXreg (FDR$\leq$ 0.25 from PA-test)
pdp btw = predicted degradation phase between

C.6 Shape of Time Series

In Section 2.7.4 I investigate the shape of time series. For that I first find for the transcriptional activity the best fit out of different lengths of Fourier series. Then I test if a fit of the transcript abundance would work better if the fitting model has one more Fourier term than the fit of the corresponding transcriptional activity. To compare fits I use the ANOVA package implemented in R.

The algorithm to find the best suitable fit for the transcriptional activity is as follows:

1. Initialize number of terms: $n = 1$, corresponds to a cosine function*
2. Initialize fake start p-value: $p_{\text{compare}} = 0.1$
3. Perform two linear fits:
   a) Fourier series with n terms
   b) Fourier series with $n + 1$ terms
4. Perform ANOVA test to compare both fits, get p-value p_{ANOVA}
5. If $\frac{p_{\text{compare}}}{p_{\text{ANOVA}}} > 10$, then best fit for transcriptional activity is a Fourier series with **order= n + 1**.
6. Else: $n = n + 1$, $p_{\text{compare}} = p_{\text{ANOVA}}$, go to 3.

* My first try is already an oscillating fit and with that I loose all constant time series. However, I feared that the step from a constant to a rhythmic fit will give highly significant test-result and any differences between fits of Fourier series with higher terms would be not as significant.

I then used the order received from the first algorithm to compare two fits for the corresponding mRNA abundance, first a Fourier series with order from transcriptional activity fit, second a Fourier series with order+1. The comparison was done using again the ANOVA test which provided the in the main text reported p-values.

To test for enrichment between both tests I used again a Fisher exact test. My contingency table I tested had the following form for each organ with x =q-value in Figure 2.14B:

	q-value$<x$	q-value$\geq x$	Total
PA-test	A	B	A+B
test for shape	C	D	C+D
Total	A+C	B+D	A+B+C+D

Appendix D: Harmonics Generated by Rhythmic Splicing

D.1

In Section 3.1 I investigate if rhythmic splicing can result in higher harmonics in the mRNA abundance, *i.e.* mRNA with a period of 12 hours instead of 24 hours.

A pre-RNA x is rhythmically transcribed and rhythmically spliced and serves as production for an mRNA y which is then degraded with a constant degradation rate. This is described by the model

$$\frac{\mathrm{d}x}{\frac{d}{t}} = \mathrm{prod}(t) - \gamma_{\mathrm{splic}}(1 + A_{\mathrm{splic}}\cos(\omega t - \phi_{\mathrm{splic}})) \tag{D.1}$$

$$\frac{\mathrm{d}y}{\frac{d}{t}} = k\gamma_{\mathrm{splic}}(1 + A_{\mathrm{splic}}\cos(\omega t - \phi_{\mathrm{splic}})) - \gamma_{\mathrm{deg}} \tag{D.2}$$

I simulated the model with the parameters, if not indicated, $\omega = \frac{2\pi}{24}\mathrm{h}^{-1}$, $\gamma_{\mathrm{splic}} = \ln 2/0.5\mathrm{h}^{-1}$, $k = 1\mathrm{h}^{-1}$, $\gamma_{\mathrm{deg}} = 0.5\mathrm{h}^{-1}$. For the function describing the transcriptional activity, I tested two periodic functions, in the main text termed cosine and "pointy" function. The equations for both read:

$$\mathrm{prod}_{\mathrm{cos}}(t) = (1 + A_{\mathrm{prod}}\cos(\omega t)) \tag{D.3}$$

$$\mathrm{prod}_{\mathrm{point}}(t) = (1 + A_{\mathrm{prod}}\frac{1}{1 + 0.8\cos(\omega t)}) \tag{D.4}$$

In Figure D.1 $\mathrm{prod}_{\mathrm{point}}(t)$ for different relative amplitudes is shown.

Detection of period was done by counting maxima of the time series within 24 hours. The simulation, analysis of simulated time series and graphical representation was done using a Mathematica script.

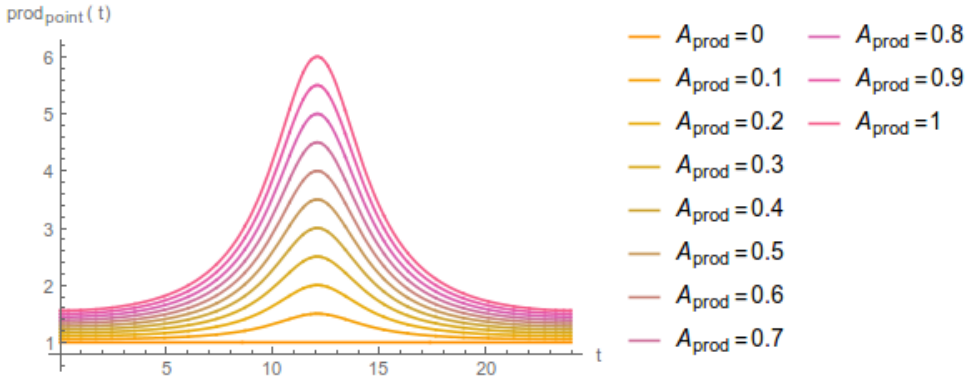


Figure D.1: The function $\text{prod}_{\text{point}}(t)$ for different relative amplitudes.

Appendix E: Rhythmic postTRXreg in Core Clock

E.1

In Section 3.2 I investigated a Goodwin model with rhythmic degradation rates. The Goodwin model I used reads:

$$\frac{\mathrm{d}x}{\frac{d}{t}} = k_x \frac{K^h}{K^h + z^h} - d_x(1 + A_x \cos(\omega t - \phi_x))x \tag{E.1}$$

$$\frac{\mathrm{d}y}{\frac{d}{t}} = k_y x - d_y(1 + A_y \cos(\omega t - \phi_y))y \tag{E.2}$$

$$\frac{\mathrm{d}z}{\frac{d}{t}} = k_z y - d_z(1 + A_z \cos(\omega t - \phi_z))z \tag{E.3}$$

For simulation I used the following parameters $k_x = k_y = k_z = 1$, $d_x = d_y = d_z = 0.1$, $K = 1$, $h = 10$. If simulated this system with constant degradation rates, that is $A_x = A_y = A_z = 0$, all variables oscillated with a period $T = 39.71$. I used this period as reference and fixed the angular frequency of the degradation rates with $\omega = \frac{2\pi}{T}$.

I simulated the system and let either one of the degradation rates oscillate, all others were set to constant, or in other words their relative amplitude was set to zero. I determined the oscillation properties of each variable by following algorithms (`x` stands for any variable of interest, x, y or z in the model equation):

period/phase:

1. Iterate through time points starting at `t=500`,
   find a peak (`x_t < x_t+1 & x_t+1 > x_t+2`).
2. Iterate further through time points, find second peak.
3. `phase = t_peak2 - t_peak1`
4. `period = t_peak2 - t_peak1`

magnitude:

1. `M = 0`
2. Iterate through time points over one period:
   For each time point `n`: `M = M + x[n]*(time[n+1]-time[n])`
3. `M = M/period`

rel. amplitude:

1. `x_max=0, x_min=1E+06`
2. Iterate through time points `n` over one period:
   For each time point `n`:
   `if x[n] < x_min:  x_min = x[n]`

   `if x[n] > x_max:  x_max = x[n]`
3. `relAmp = 1/M * 0.5 * (x_max-x_min)`

Simulation, time series analysis and graphical representation was done using a python script.

Appendix F: Summary Tables

F.1 List of Circadian Core Clock Genes

List of circadian genes used in Figure 2.3 (MGI-symbols):
Arntl, Npas2, Cry1, Cry2, Per1, Per2, Per3, Dbp, Nr1d1, Nr1d2

F.2 Summary Table of Data Analysis

	liver	kidney
expressed genes: TRXact & mRNA	12940	13711
TRXact: circadian	3028	4591
mRNA: circadian	3813	4137
circadian rhythm in TRXact or mRNA ("circadian genes")	5581	6489
genes expressed in both organs	12148	
circadian genes in both organs (% of 12148)	1361 (11%)	
genes with half-life either measured by		
Friedel *et al.* [212] or Schwanhäusser *et al.* [213]	6949	6983
genes with half-life measured in both publications	3103	3097
Test from Section 2.3, 1^{st} stage, genes outside of semi-circle		
(% of tested genes = all circadian)	2191 (∼36%)	939 (∼14%)
2^{nd} stage, tested genes (within semi-circle + half-life measured at least once)	850	1163
2^{nd} stage, genes with no agreement with model prediction		
(% of tested genes)	483 (∼42%)	419 (∼49%)
Test from Section 2.4.1, PA-test [179], tested genes (% of circadian genes)	1023	899
PA-test [179], positively tested for rhythmic postTRXreg (% of tested genes)	349 (34%)	899 (18%)
Test from Section 2.4.2, tested all circadian genes (no half-life required),		
positively tested for rhythmic postTRXreg	52	11

Table 1: Abbreviation: TRXact = transcriptional activity

F.3 List of Genes with Rhythmic Post-Transcriptional Control

List of genes with rhythmic post-transcriptional regulation found by the pa-test with FDR≤ 0.25 in Section 2.4. If the gene was found in liver or kidney the predicted degradation phase ϕ_{deg} is shown.

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
1	1190002N15Rik	2.09	
2	2410002F23Rik	15.83	
3	2510039O18Rik	13.98	
4	Aco1		20.45
5	Aco2	23.08	18.23
6	Acot9	19.87	
7	Acsl5	7.99	
8	Actn1	22.64	13.43
9	Adar		11.05
10	Adk		20.82
11	Aebp1	17.36	
12	Aes		23.32
13	Agpat4		17.79
14	Agps	15.83	
15	Ahctf1		0.81
16	Ahcy		18.33
17	AI837181	18.92	
18	Ak2		15.16
19	Akap11		10.99
20	Akr1b8		17.69
21	Alad		22.18
22	Aldh9a1		21.62
23	Alg2	3.01	
24	Amacr		8.66
25	Angptl2	19.47	17.13
26	Ank3		14.76
27	Ankrd12		18.76
28	Anp32b	17.41	
29	Anxa11		16.57
30	Anxa2	2.24	
31	Anxa3		15.31
32	Anxa7	18.82	14.54
33	Ap1m1	22.27	
34	Ap2a2	4.58	2.80
35	Ap2s1		16.47
36	Aprt		8.86
37	Arhgap28	3.30	
38	Arl8a		12.65
39	Asna1	21.77	
40	Atad2		20.27
41	Atp6ap1		1.46
42	Atp6v1e1		13.31
43	B230219D22Rik		12.64
44	Bag2	23.94	
45	Bag3		21.16
46	Baz1b		18.13
47	BC017643	7.20	
48	Becn1	21.43	
49	Bgn		19.09
50	Bicd2	1.51	

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
51	Bid	3.93	
52	Bloc1s2		14.50
53	Cacybp	0.22	
54	Cald1	6.60	5.25
55	Cand2	7.05	
56	Cask		19.97
57	Casp3		13.17
58	Cbl	20.83	
59	Cbx6		15.92
60	Cct3	23.77	
61	Cct5	0.96	23.96
62	Cd2ap		9.36
63	Cd3eap	18.56	
64	Cd9		7.47
65	Cdca8		12.56
66	Cdh11	8.01	
67	Cdk5rap2		10.86
68	Cdkn2c		10.42
69	Chaf1b		15.74
70	Chchd4		15.71
71	Chka		3.42
72	Chmp2a		17.09
73	Chmp6		7.33
74	Chordc1	1.76	22.76
75	Cirbp	13.87	9.11
76	Clic4	3.63	0.09
77	Clk1	5.75	
78	Clpp	19.68	
79	Clpx	7.39	2.85
80	Cltb	23.02	
81	Cmtm7	10.23	
82	Cobll1		14.87
83	Coil	4.32	
84	Col6a1	23.01	10.69
85	Comt	7.64	
86	Cpne8		8.39
87	Cpsf2		5.45
88	Cpt1a	16.43	15.48
89	Crcp		18.12
90	Creld2	23.49	
91	Csad	19.25	
92	Cstf2t		13.43
93	Cttn		12.01
94	Cttnbp2nl		16.61
95	Cyc1		20.43
96	Cyr61		7.68
97	D1Ertd622e		17.47
98	Dab2	15.29	18.91
99	Dap3		15.46

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
100	Dazap1		13.23
101	Dbi		1.00
102	Dbn1		19.90
103	Dbr1	5.68	
104	Dcakd		0.72
105	Ddx20	20.64	
106	Ddx47		10.08
107	Ddx5	18.64	
108	Ddx52		7.79
109	Dhfr		10.68
110	Dlst		18.00
111	Dnajb1		22.29
112	Dnajb6		16.52
113	Dnajc1		21.86
114	Dpy19l1		23.75
115	Dus3l	21.12	
116	Dusp14		22.87
117	Dusp3		17.90
118	Dynll1	1.46	
119	Ecm1		18.79
120	Eef1e1	2.57	
121	Efhd2		16.99
122	Ehbp1		5.28
123	Ehd1	1.99	
124	Eif4b		12.26
125	Elavl1		4.49
126	Emb		8.67
127	Emg1	23.57	
128	Eml3		4.62
129	Eml4		0.99
130	Esd		0.02
131	Etfa		21.91
132	Evi5		2.76
133	Exoc3		9.87
134	Extl2		12.12
135	F3		10.73
136	Fadd		11.30
137	Faim		8.92
138	Fbxo21	10.69	10.83
139	Fbxo22		12.34
140	Fbxo4		8.74
141	Fdxr		16.04
142	Fis1		16.78
143	Fkbp10	17.36	
144	Fkbp4	1.36	23.24
145	Fkbp5		18.71
146	Flot1		11.23
147	Foxk1		13.25
148	Foxk2	1.12	
149	Fpgs		0.69

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
150	Fundc2		9.89
151	Fyn		14.18
152	Gclc		3.05
153	Gclm		4.15
154	Gga2		21.44
155	Glrx		18.05
156	Glud1		0.18
157	Glul	21.21	
158	Gne		14.02
159	Gng12		13.24
160	Gnpnat1		10.85
161	Gns	20.66	15.29
162	Golgb1		12.52
163	Gopc	20.33	
164	Gpd1l		17.10
165	Gsk3a	22.63	17.36
166	Gsk3b		21.58
167	Gsto1		12.06
168	Gtf2h5		20.93
169	H2-D1		17.36
170	Hdgf	0.69	17.99
171	Hdgfrp3		10.55
172	Higd2a	0.91	
173	Hmg20a		12.47
174	Hmgb2	11.55	10.40
175	Hmgb3	12.48	
176	Hn1		8.36
177	Hpcal1		17.11
178	Hs1bp3		11.21
179	Hsd17b7	21.34	
180	Hsp90aa1		22.61
181	Hsp90ab1		22.01
182	Hsp90b1	1.55	
183	Hspa14		15.43
184	Hspa4l	1.09	22.72
185	Hspa8		21.39
186	Hspb8		17.36
187	Ifit1		11.95
188	Igf2r		9.48
189	Igfbp7	20.88	
190	Ilf2		11.02
191	Immt		14.66
192	Ints3	1.65	
193	Iqgap1		12.75
194	Itpr1		19.60
195	Ivns1abp		16.78
196	Jmjd1c		2.93
197	Kars		9.98
198	Kdelr2	21.81	
199	Khdrbs1		11.24
200	Klf13		17.45

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
201	Kpna1	2.89	
202	Kpnb1		1.79
203	Lamp2		3.54
204	Lasp1		4.88
205	Leo1	3.66	1.29
206	Lgals3		11.80
207	Lima1		13.17
208	Lman2	1.57	
209	Lmnb2	2.93	
210	Lmo7		17.21
211	Lrrc8d		7.12
212	Luc7l2	8.58	
213	Mad2l1bp	3.01	
214	Man2a1		18.69
215	Maoa	12.13	
216	Map2k7		10.29
217	Mapk14	21.85	19.95
218	Mapre2		19.40
219	Marcks		11.70
220	Mbnl2		18.39
221	Mcc	1.70	
222	Mcfd2	22.86	22.07
223	Mcm2		18.67
224	Mdh2		19.12
225	Med19		8.53
226	Med28		7.80
227	Med8		3.30
228	Metap1		0.38
229	Mical2	6.15	2.47
230	Mnat1		12.99
231	Mocs3		13.82
232	Morc3		19.78
233	Mrpl22		8.90
234	Mrpl32		10.20
235	Mrpl50	11.11	
236	Mrpl9	17.42	
237	Mrps18b	17.86	
238	Mrps24		12.80
239	Mrps26	16.58	
240	Mrps30	5.48	
241	Mrps34		12.72
242	Mrps5		13.26
243	Msh6		11.48
244	Msn		12.93
245	Mtap		22.16
246	Mtx1	15.74	
247	Myadm		16.29
248	Myo1b		1.13
249	Myo5a	18.70	
250	Napa		16.62

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
251	Nat10		1.63
252	Ncoa5		3.24
253	Ndufa2		20.11
254	Nedd4l		1.75
255	Nek6		6.17
256	Nin	11.13	
257	Nisch		18.89
258	Noc3l	5.18	
259	Npc2		13.94
260	Npr3		17.65
261	Nt5c	17.13	
262	Nt5c2		16.77
263	Nubp2	19.28	
264	Nucb1		8.31
265	Nudcd3		16.10
266	Nudt16l1		14.77
267	Nudt3		16.96
268	Nup62		0.31
269	Nup93		9.60
270	Ogfod1		19.60
271	Ormdl3		23.47
272	Osbpl11		1.95
273	Oxnad1		19.73
274	P4ha2	1.89	21.66
275	Pak1ip1	12.13	
276	Palld		0.69
277	Pcbp4	17.47	12.24
278	Pdcl		10.26
279	Pdia6	1.18	
280	Pdss2		4.91
281	Pdxk		17.71
282	Pebp1		21.21
283	Pex16	14.38	
284	Pfdn2		15.19
285	Pfn1	4.24	
286	Pgk1		1.90
287	Phb		19.11
288	Pik3r1		19.96
289	Plk1		13.53
290	Plod2	8.09	
291	Plscr1		9.64
292	Pmpca	20.92	
293	Pmvk		21.98
294	Pnp	8.30	23.61
295	Pold3		14.04
296	Pole3		14.44
297	Polr2b		19.06
298	Polr2d		8.74
299	Polr3g		23.91

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
300	Pon2		13.37
301	Pop4		11.70
302	Pop5		2.35
303	Por		16.99
304	Ppat		15.60
305	Ppfibp1	11.84	8.46
306	Ppm1l		20.79
307	Ppp1ca		2.81
308	Ppp1r14b	20.22	
309	Ppp2r1a		7.58
310	Ppp2r1b	4.40	
311	Ppp2r5e		20.04
312	Pptc7		16.20
313	Prepl		6.49
314	Prkar2a		23.00
315	Prkcd		12.48
316	Prkci		1.02
317	Prmt6	6.10	
318	Prpf38a		9.32
319	Prpf38b	5.00	
320	Prps1	5.50	
321	Psen1	20.97	
322	Psip1	5.28	
323	Psma4		10.23
324	Psmc2		3.92
325	Psmc5	22.90	
326	Psmd10		8.05
327	Psmd11		5.45
328	Psmd2		8.06
329	Psmd7	0.60	
330	Psmd9		14.53
331	Psme3	22.56	
332	Ptpn1		13.81
333	Ptpn12	1.76	
334	Pura	5.86	
335	Pus7		17.58
336	Qtrtd1		13.58
337	Rab1b		11.85
338	Rab21		3.33
339	Rab34	18.08	8.54
340	Rabep1		2.78
341	Rad50	14.02	5.87
342	Rai14		4.39
343	Rapgef1		15.57
344	Rapgef6		9.10
345	Rasa3	14.21	14.50
346	Rbm14		17.77
347	Rbm17		12.31
348	Rbm3		20.71
349	Rbm7		13.23
350	Rbms1		13.07

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
351	Rcl1	9.78	12.21
352	Rcor1	3.52	
353	Reps1		10.64
354	Rfx1	1.03	
355	Rhob		16.74
356	Rhoc		10.84
357	Rpl26	22.44	
358	Rpl32		20.41
359	Rps15	21.73	
360	Rps25	18.14	
361	Rps27a		2.57
362	Rps29		14.41
363	Rps4x	19.72	
364	Rps6ka2		14.10
365	Rps6ka4		13.43
366	Rpusd4		13.84
367	Rragc		13.90
368	Rrm2		18.79
369	Rtkn	20.31	
370	Samhd1		13.70
371	Scd2	1.31	
372	Sdc1		23.86
373	Sec23b	20.68	
374	Sephs1		9.59
375	Sept8		16.98
376	Setd7	4.91	
377	Sf3b2	19.16	
378	Sf3b5		15.42
379	Sfxn1		17.29
380	Sgpl1		2.42
381	Shmt1	7.06	
382	Shmt2		17.42
383	Slc12a4		12.81
384	Slc16a1		3.20
385	Slc1a4		5.67
386	Slc25a13		19.38
387	Slc25a20	21.65	
388	Slc25a24		11.20
389	Slc25a32		15.97
390	Slc25a4		17.16
391	Slc44a2		12.48
392	Slc6a6	14.41	3.29
393	Smarcd2		18.01
394	Smarce1		6.07
395	Snw1		13.42
396	Snx12	4.01	
397	Snx3	0.32	22.91
398	Sod2		17.67
399	Spag5	22.54	

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
400	Spag7		14.79
401	Spast		2.84
402	Sra1		18.53
403	Srgap3		3.16
404	Stip1	0.50	23.82
405	Stx18	2.07	
406	Stx5a	20.72	
407	Syngr2		16.15
408	Taf9	19.12	10.02
409	Tars		3.09
410	Tbc1d15		11.07
411	Tbrg4	17.70	
412	Tceb3		3.82
413	Tcp1	22.76	23.60
414	Tex10		7.17
415	Tex2		2.98
416	Timm13	18.49	
417	Timm17a		12.19
418	Tmed1	10.74	
419	Tmem109	19.81	
420	Tmem50b	23.69	
421	Tmem97		16.69
422	Top2a		11.66
423	Tor1a		17.19
424	Tor1b		15.86
425	Tor3a	10.78	
426	Trappc3		12.97
427	Trappc4		11.66
428	Trappc5		21.20
429	Trim24		17.66
430	Trim65		20.91
431	Trip10	15.72	
432	Trove2		2.48
433	Tsc22d1	23.91	5.83
434	Tspan4	19.75	20.56
435	Tubb2b	3.83	4.10
436	Tubb6	6.31	
437	Tubg1		22.19
438	Txnrd3		9.63
439	Ube2z	0.47	
440	Ubl3		17.19
441	Ubqln1	2.98	2.91
442	Ubr2		17.95
443	Ubtf		13.77
444	Upf1		19.77
445	Utp14a		8.75
446	Utp15		12.52
447	Utp6		8.72
448	Utrn		12.57
449	Vamp2	16.65	16.48
450	Vamp3		15.74

	MGI	predicted ϕ_{deg} Kidney	predicted ϕ_{deg} Liver
451	Vcam1		17.41
452	Vps13a	23.26	19.33
453	Vps37c		14.51
454	Wbp2		14.98
455	Wdr33		6.01
456	Wdr6		9.94
457	Wrnip1		2.10
458	Ypel5		13.91
459	Ywhag		19.75
460	Zcchc11		14.79
461	Zcrb1		21.02
462	Zfr		23.35
463	Zmym4		9.27

F.4 RNA Binding Proteins with Rhythmic Transcript

All gene functions were taken either from OMIM (`www.omim.org`), Wikipedia (`wikipedia.org`) or direct publications abstracts, indicated by **O**, **W** or **P**. I tried to find gene functions associated with the RNA-binding. However, this was not always possible.

	MGI	Organ	Function
1	A1cf	kid	APOBEC1 complementation factor Mammalian apolipoprotein B (APOBEC1 = A1cf) mRNA undergoes site-specific C to U deamination, which is mediated by a multicomponent enzyme complex containing a minimal core composed of APOBEC1 and a complementation factor encoded by this gene [267]. The gene product has three non-identical RNA recognition motifs and belongs to the hnRNP R family of RNA-binding proteins. It has been proposed that this complementation factor functions as an RNA-binding subunit and docks APOBEC1 to deaminate the upstream cytidine. Studies suggest that the protein may also be involved in other RNA editing or RNA processing events. Its deletion results in lethality in mice. [268] **W**
2	Aptx	liv	Aprataxin associated with DNA single-strand breaks and their repair, see OMIM entry for more details. APTX efficiently repairs adenylated RNA-DNA, and acting in an RNA-DNA damage response, promotes cellular survival and prevents S-phase checkpoint activation in budding yeast undergoing RER. Structure-function studies of human APTX-RNA-DNA-AMP-zinc complexes define a mechanism for detecting and reversing adenylation at RNA-DNA junctions. This involves A-form RNA binding, proper protein folding, and conformational changes, all of which are affected by heritable APTX mutations in ataxia with oculomotor apraxia-1. Tumbale *et al.* [269] concluded that accumulation of adenylated RNA-DNA may contribute to neurologic disease. Garcia-Diaz *et al.* [270] found that most, but not all, cell lines derived from AOA1 patient fibroblasts showed coenzyme Q10 (CoQ10) deficiency due to reduced mRNA and protein expression of PDSS1, the first committed enzyme of CoQ10 biosynthesis. Low PDSS1 was caused by reduced activity of a transcriptional regulatory pathway that included APE1, NRF1, and NRF2. Knockdown of APTX or APE1 in HeLa cells recapitulated CoQ10 deficiency and other mitochondrial abnormalities, and these abnormalities were reversed by upregulation of NRF2. Garcia-Diaz *et al.* [270] concluded that mitochondrial dysfunction in APTX-depleted cells is not due to involvement of APTX in mtDNA repair, but rather to a role for APTX in transcriptional regulation of mitochondrial function. **O**
3	Cirbp	liv/kid	Cold inducible RNA binding protein Morf *et al.* [168] showed that simulated body temperature cycles, but not peripheral oscillators, controlled the rhythmic expression of cold-inducible RNA-binding protein (CIRBP) in cultured fibroblasts. In turn, loss-of-function experiments indicated that CIRBP was required for high-amplitude circadian gene expression. The transcriptomewide identification of CIRBP-bound RNAs by a biotin-streptavidin-based crosslinking and immunoprecipitation (CLIP) procedure revealed several transcripts encoding circadian oscillator proteins, including CLOCK. Moreover, CLOCK accumulation was strongly reduced in CIRBP-depleted fibroblasts. Because ectopic expression of CLOCK improved circadian gene expression in these cells, Morf *et al.* [168] concluded that CIRBP confers robustness to circadian oscillators through regulation of CLOCK expression. **O**

	MGI	Organ	Function
4	Cpeb1	liv	Cytoplasmic polyadenylation element binding protein 1 CPEB1 binds the cytoplasmic polyadenylation element (CPE), a uridine-rich sequence element within the mRNA 3-prime untranslated region (UTR) involved in directing cytoplasmic polyadenylation and translational activation. CPEB1 has been implicated in mediating both polyadenylation-dependent translation activation and CPE-directed translational repression [271, 272] **O**
5	Cpeb3	liv/kid	Cytoplasmic polyadenylation element binding protein 3 Using an in vitro self-selection technique, Salehi-Ashtiani *et al.* [273] identified self-cleaving ribozymes associated with olfactory receptor OR4K15, insulin-like growth factor-1 receptor (IGF1R), a LINE-1 retrotransposon (LRE1), and CPEB3. **O**
6	Cpeb4	liv	Cytoplasmic polyadenylation element binding protein 4 Cytoplasmic polyadenylation elements (CPEs) are located in the 3-prime UTRs of hundreds of mRNAs involved in cell proliferation, chromosome segregation, and cell differentiation. CPEs function as docking sites for CPE-binding proteins (CPEBs), which can either repress or activate mRNA translation. CPEB4 activates mRNA translation by adding poly(A) tails to target mRNAs [274]**O**
7	Cpsf4l	kid	Cleavage and polyadenylation specific factor 4 like
8	Csdc2	kid	Cold shock domain-containing C2
9	Dnajc17	kid	DnaJ heat shock protein family (Hsp40) member C17 Members of the Heat-shock 40-kd protein family, such as DNAJC17, are involved in transcriptional regulation [275]. **O**
10	Enox1	kid	ecto-NOX disulfide-thiol exchanger 1 Electron transport pathways are generally associated with mitochondrial membranes, but non-mitochondrial pathways are also biologically significant. Plasma membrane electron transport pathways are involved in functions as diverse as cellular defense, intracellular redox homeostasis, and control of cell growth and survival. Members of the ecto-NOX family, such as CNOX, or ENOX1, are involved in plasma membrane transport pathways. These enzymes exhibit both a hydroquinone (NADH) oxidase activity and a protein disulfide-thiol interchange activity in series, with each activity cycling every 22 to 26 minutes [276] **O**
11	Fus	liv/kid	FUS RNA binding protein FUS is a nucleoprotein that functions in DNA and RNA metabolism, including DNA repair, and the regulation of transcription, RNA splicing, and export to the cytoplasm. Translocation of the FUS transcriptional activation domain results in fusion proteins and has been implicated in tumorigenesis (summary by Vance *et al.* [277])

	MGI	Organ	Function
12	Igf2bp2	liv/kid	Insulin-like grwoth factor2 mRNA-binding protein 2 The Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes for BioMedical Research analyzed 386,731 common SNPs in 1,464 patients with type 2 diabetes (see 125853) and 1,467 matched controls, each characterized for measures of glucose metabolism, lipids, obesity, and blood pressure. With collaborators Finland-United States Investigation of NIDDM Genetics (FUSION) and Wellcome Trust Case Control Consortium/United Kingdom Type 2 Diabetes Genetics Consortium (WTCCC/UKT2D), this group identified and confirmed a locus associated with type 2 diabetes in the second intron of IGF2BP2. [278] **O**
13	Khdrbs3	kid	KH RNA binding domain containing, signal transduction associated 3 KHDRBS3 is an RNA-binding protein involved in regulation of alternative splicing [279]. **O**
14	Lsm3	liv	Lsm3 protein Sm-like proteins were identified in a variety of organisms based on sequence homology with the Sm protein family. Sm-like proteins contain the Sm sequence motif, which consists of 2 regions separated by a linker of variable length that folds as a loop. The Sm-like proteins are thought to form a stable heteromer present in tri-snRNP particles, which are important for pre-mRNA splicing. **O**
15	Lsm8	liv	see Lsm3
16	Mbnl2	liv	Muscleblind-like splicing regulator 2 MBNL2 belongs to a conserved family of RNA-binding proteins characterized by 2 pairs of C3H-type zinc finger-related motifs. These proteins function as target-specific regulators of pre-mRNA splicing [280]. **O**
17	Parn	liv	Poly(A)-specific ribonuclease deadenylating nuclease The PARN gene, which belongs to a highly conserved family of exoribonucleases, acts by shortening mRNA poly(A) tail length through the process of deadenylation, thus regulating gene expression [166]. **O**
18	Parp14	liv	Poly(ADP-ribose) polymerase 14 Poly(ADP-ribosyl)ation is an immediate DNA damage-dependent posttranslational modification of histones and other nuclear proteins that contributes to the survival of injured proliferating cells. PARP14 belongs to the superfamily of enzymes that perform this modification [281]. **O**
19	Pcbp3	kid	Poly(rC)-binding protein 3 The poly(rC)-binding proteins are a subfamily of KH-domain-containing RNA-binding proteins that bind C-rich pyrimidine tracts and play pivotal roles in a broad spectrum of posttranscriptional events [282]. **O**

	MGI	Organ	Function
20	Pcbp4	liv	Poly(rC)-binding protein 4, see Pcbp3
21	Ppargc1a	liv	Peroxisome proliferator-activated receptor-gamma, coactivator 1 PPARGC1A is a coactivator of nuclear receptors and other transcription factors that regulate metabolic processes, including mitochondrial biogenesis and respiration, hepatic gluconeogenesis, and muscle fiber-type switching [283]. The PPARGC1 protein contains a putative RNA-binding domain and 2 SR domains. Proteins containing paired RNA-binding motifs and SR domains interact with the C-terminal domain of RNA polymerase-2. **O**
22	Pspc1	liv	Paraspeckle component 1 Fox *et al.* [284] found that PSPC1, PSPC2, and p54NRB relocalized from paraspeckles to the perinucleolar cap region upon transcriptional blockade. **O**
23	Ralyl	kid	Raly-like protein
24	Rbfox1	kid	RNA-binding protein Fox1 Rbfox1 regulates tissue-specific splicing by binding to the element (U)GCAUG in mRNA precursors. Depending on where it binds relative to the regulated exon, Rbfox1 can regulate splicing positively or negatively [285]. **O**
25	Rbm11	kid	RNA-binding motif protein 11 Tissue-specific splicing regulator. RBM11 is selectively expressed in brain, cerebellum and testis, and to a lower extent in kidney. [286] **P**
26	Rbm12b1	liv	RNA-binding motif protein 12B1
27	Rbm25	liv	RNA-binding motif protein 25 Rbm25 associated with multiple splicing components, including CBP80 (NCBP1), Sm proteins, and small nuclear RNAs. Rbm25 maintained its association with mRNA after splicing, and it associated with an mRNA-derived exon product following splicing more efficiently than a cDNA-derived exon. [287]**O**
28	Rbm38	kid	RNA-binding motif protein 38 RNPC1 expression was induced by DNA-damaging agents in cells expressing wildtype p53, but not in cells lacking p53 expression [288]. Overexpression of RNPC1a, but not RNPC1b, induced cell cycle arrest in G1, and cell cycle arrest was independent of p53 or expression of the p53 target protein p21. Both RNPC1a and RNPC1b bound the 3-prime region of the p21 transcript, but only RNPC1a increased p21 mRNA and protein levels. Northern blot analysis revealed that the half-life of p21 mRNA was more than doubled by RNPC1a expression. Shu *et al.* [288] concluded that RNPC1a mediates p53-induced cell cycle arrest by stabilizing p21 mRNA. **O**

29	Rbm46	kid	RNA-binding motif protein 46 Blastocyst formation represents the first lineage specification by segregation of the trophectoderm from the inner cell mass in early embryonic development. Rbm46 regulates trophectoderm differentiation by stabilizing Cdx2 mRNA in early mouse embryos [289]. **P**
30	Rbm47	liv	RNA-binding motif protein 47 Alternative splicing (AS) plays a critical role in cell fate transitions, development, and disease. Recent studies have shown that AS also influences pluripotency and somatic cell reprogramming. Cieply *et al.* [290] profiled transcriptome-wide AS changes that occur during reprogramming of fibroblasts to pluripotency. This analysis revealed distinct phases of AS, including a splicing program that is unique to transgene-independent induced pluripotent stem cells (iPSCs). Changes in the expression of AS factors Zcchc24, Esrp1, Mbnl1/2, and Rbm47 were demonstrated to contribute to phase-specific AS. **P**
31	Rbpms	liv	RNA-binding protein gene with multiple splicing Using yeast 2-hybrid screens, coaffinity purification analysis of transfected HEK293 cells, and bioinformatic analysis, Lim *et al.* [291] developed an interaction network for 54 human proteins involved in 23 inherited ataxias. By database analysis, they expanded the core network to include more distantly related interacting proteins that could function as genetic modifiers. ATXN1 (showed a strong direct interaction with RBPMS, and the N-terminal portion of ATXN1 was required for interaction with RBPMS. RBPMS was a main hub in the network and interacted with many proteins, including 2 cerebellar ataxia-associated proteins, ATN1 and QK1. **O**
32	Recql4	liv	Recq protein-like 4 RECQL4 from HeLa cells was isolated as a stable complex with UBR1 and UBR2, which are ubiquitin ligases of the N-end rule pathway, Although the known role of UBR1 and UBR2 is to mediate polyubiquitylation (and subsequent degradation) of their substrates, the UBR1/2-bound RECQL4 was not ubiquitylated in vivo and was a long-lived protein in HeLa cells [292]. RECQL4 siRNA knockdown human fibroblasts accumulated more H2O2-induced DNA strand breaks than control cells, suggesting that RECQL4 may stimulate repair of H2O2-induced DNA damage [293]. **O**
33	Sugp2	liv	splicing factor, Arginine/Serine-rich 14 (SFRS14) SFRS14 is a member of the SURP family, proteins containing the RNA binding motif SURP, of splicing factors. **O**
34	Synj2	liv	Synaptojanin 2 Chuang *et al.* [294] found that small interfering RNA-mediated depletion of RAC1 or SYNJ2 in 2 human glioblastoma cell lines inhibited migration of the cells through 3-dimensional gel and rat brain slices, and it inhibited cell migration on glioma-derived extracellular matrix. **O**

35	Tab2	liv	TAK1-binding protein 2 connected to NfκB regulation, see OMIM entry for more details
36	Thumpd2	kid	Thump domain-containing protein 2
37	Tra2a	kid	Transformer 2 alpha Tacke *et al.* [295] showed that human TRA2A and TRA2B proteins are present in HeLa cell nuclear extracts and that they bind efficiently and specifically to a previously characterized pre-mRNA splicing enhancer element. Both purified proteins bind preferentially to RNA sequences containing GAA repeats, characteristic of many enhancer elements. Neither TRA2 protein functions in constitutive splicing in vitro, but both activate enhancer-dependent splicing in a sequence-specific manner and restore it after inhibition with competitor RNA. These findings indicate that mammalian TRA2 proteins are sequence-specific splicing activators that are likely to participate in the control of cell-specific splicing patterns. **O**
38	Trnau1ap	liv	tRNA selenocysteine 1 associated protein 1
39	Trove2	liv	Trove domain family, member 2 Hung *et al.* [296] cataloged the Trove2-associated RNAs in human cell lines and found that among other RNAs, Trove2 bound an RNA motif derived from endogenous Alu retroelements. Alu transcripts were induced by type I interferon and stimulated proinflammatory cytokine secretion by human peripheral blood cells. Trove2 deletion resulted in enhanced expression of Alu RNAs and interferon-regulated genes. **O**
40	Zc3h12a	liv	Zinc finger CCCH domain-containing protein 12A ZC3H12A is a CCCH-type zinc finger protein, whereas most mammalian zinc finger proteins are CCHH- or CCCC-type proteins. The ZC3H12A gene encodes an essential RNase that controls the stability of a set of inflammatory genes [297]. **O**
41	Zc3h12d	liv/kid	Zinc finger CCCH domain-containing protein 12D Liang *et al.* ([298] found that ZC3H12B, also known as MCPIP2, and other MCPIP proteins, MCPIP1 (ZC3H12A), MCPIP2 (ZC3H12B), and MCPIP3 (ZC3H12C), regulate macrophage activation. **O**
42	Zc3h6	liv/kid	Zinc finger CCCH domain-containing protein 6

43	Zc3hav1l	liv	Zinc finger CCCH-type containing, antiviral 1 like Interchromosomal chimeric RNA molecules are often transcription products from genomic rearrangement in cancerous cells. Fang *et al.* [299] report the computational detection of an interchromosomal RNA fusion between ZC3HAV1L and CHMP1A from RNA-seq data of normal human mammary epithelial cells, and experimental confirmation of the chimeric transcript in multiple human cells and tissues. Our experimental characterization also detected three variants of the ZC3HAV1L-CHMP1A chimeric RNA, suggesting that these genes are involved in complex splicing. The fusion sequence at the novel exon-exon boundary, and the absence of corresponding DNA rearrangement suggest that this chimeric RNA is likely produced by trans-splicing in human cells. **P**
44	Zcchc11	liv	Zinc finger CCCH domain-containing protein 11 Minoda *et al.* [300] showed that a substantial proportion of ZCCHC11 translocated from the nucleus to the cytoplasm following lipopolysaccharide (LPS) stimulation of HEK293T cells. ZCCHC11L coimmunoprecipitated with TIFA only following LPS treatment, and it specifically suppressed LPS-induced NF-kappaB activation. Jones *et al.* [301] found that recombinant mouse Zcchc11 and endogenous human ZCCHC11 showed nucleotidyltransferase activity. ZCCHC11 did not show RNA substrate specificity and used all RNA substrates examined. **O**
45	Zcrb1	liv	Zinc finger CCCH domain- and RNA-bindng motif-containing protein 1 In the first step of U12-type spliceosome formation, U11 and U12 small nuclear ribonucleoproteins (snRNPs) bind U12-type pre-mRNAs as a preformed U11/U12 di-snRNP. The U11 and U12 components recognize the 5-prime splice site and the branch site, respectively, and together form a molecular bridge connecting both ends of the intron. ZCRB1 is a component of the U11/U12 di-snRNP [302]. **O**
46	Zfp36	liv	Zinc finger protein 36-like Galloway *et al.* [303] demonstrated in developing B lymphocytes, the RNA-binding proteins ZFP36L1 and ZFP36L2 are critical for maintaining quiescence before precursor B cell receptor (pre-BCR) expression and for reestablishing quiescence after pre-BCR-induced expansion. These RBPs suppress an evolutionarily conserved posttranscriptional regulon consisting of mRNAs whose protein products cooperatively promote transition in the S phase of the cell cycle. **O**

Bibliography

[1] M. Beuse, R. Bartling, A. Kopmann, H. Diekmann, and M. Thoma. Effect of the dilution rate on the mode of oscillation in continuous cultures of *Saccharomyces cerevisiae*. *Journal of Biotechnology*, 61(1):15–31, 1998.

[2] Susan S. Golden and Shannon R. Canales. Cyanobacterial circadian clocks timing is everything. *Nature Reviews Microbiology*, 1(3):191–199, 2003.

[3] Jay C Dunlap. Molecular Bases for Circadian Clocks. *Cell*, 96(2):271–290, 1999.

[4] Pittindrigh. Biological Rhythms. Handbook of Behavioral Neurobiology. In *The Quarterly Review of Biology*, volume 4, pages 221–221. Jürgen Aschoff, 1982.

[5] J. S. Menet, J. Rodriguez, K. C. Abruzzi, and M. Rosbash. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife*, 1(0):e00011–e00011, 2012.

[6] Gwendal Le Martelot, Donatella Canella, Laura Symul, Eugenia Migliavacca, Federica Gilardi, Robin Liechti, Olivier Martin, Keith Harshman, Mauro Delorenzi, Beatrice Desvergne, Winship Herr, Bart Deplancke, Ueli Schibler, Jacques Rougemont, Nicolas Guex, Nouria Hernandez, Felix Naef, and the CycliX consortium. Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles. *PLoS Biology*, 10(11):e1001442, 2012.

[7] De Mairan. Observation botanique. *e Hist. Acad. Imprimerie de Du Pont (Paris)*, 1729.

[8] Sutherland Simpson and J. J. Galbraith. Observations on the Normal Temperature of the Monkey and its Diurnal Variation, and on the Effect of Changes in the Daily Routine on this Variation. *Transactions of the Royal Society of Edinburgh*, 45(01):65–104, 1906.

[9] J.S. Szymanski. Die Verteilung der Ruhe- und Aktivitätsperioden bei weißen Ratten und Tanzmäusen. *Pflügers Archiv für die Gesamte Physiologie des Menschen und der Tiere*, 1918.

[10] C. P. Richter. A Behavioristic Study of the Activity of the Rat. *Comparative Psychology Monographs*, 1(2):56, 1922.

[11] M. S. Johnson. Activity and Distribution of Certain Wild Mice in Relation to Biotic Communities. *Journal of Mammalogy*, 7(4):245, 1926.

[12] Ingeborg Beling. Über das Zeitgedächtnis der Bienen. *Zeitschrift für Vergleichende Physiologie*, 9(2):259–338, 1929.

[13] Gustav Kramer. EXPERIMENTS ON BIRD ORIENTATION *. *Ibis*, 94:265–285, 1952.

[14] J. Woodland Hastings and Beatrice M. Sweeney. A persistent diurnal Rhythm of Luminescence in \emph{Gonyaulax Polyedra}. *The Biological Bulletin*, 115(3):440–458, 1958.

[15] Jürgen Von Aschoff and Rütger Wever. Spontanperiodik des Menschen bei Ausschluß aller Zeitgeber. *Naturwissenschaften*, 49(15):337–342, 1962.

[16] D. E. Bianchi. An Endogenous Circadian Rhythm in Neurospora crassa. *Journal of General Microbiology*, 35(3):437–445, 1964.

[17] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 323(6090):720–722, 1986.

[18] E Bünning. Zur Kenntnis der erblichen Tagesperioizität bei den Primärblättern von Phaseolus multiflorus. *Jahrb. Wis.*, 1935.

[19] R. J. Konopka and S. Benzer. Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 68(9):2112–2116, 1971.

[20] P. Reddy, W. A. Zehring, D. A. Wheeler, V. Pirrotta, C. Hadfield, J. C. Hall, and M. Rosbash. Molecular analysis of the period locus in Drosophila melanogaster and identification of a transcript involved in biological rhythms. *Cell*, 38(3):701–710, 1984.

[21] M. H. Vitaterna, D. P. King, A. M. Chang, J. M. Kornhauser, P. L. Lowrey, J. D. McDonald, W. F. Dove, L. H. Pinto, F. W. Turek, and J. S. Takahashi. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science*, 264(5159):719–725, 1994.

[22] Roberto Refinetti. *Circadian physiology*. Taylor & Francis Inc, 2016.

[23] V. Dvornyk, O. Vinogradova, and E. Nevo. Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2495–2500, 2003.

[24] Mark A. Woelfle, Yao Xu, Ximing Qin, and Carl Hirschie Johnson. Circadian rhythms of superhelical status of DNA in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47):18819–18824, 2007.

[25] Vikram Vijayan, Rick Zuzow, and Erin K. O'Shea. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22564–22568, 2009.

[26] Stanly B. Williams. A Circadian Timing Mechanism in the Cyanobacteria. In *Advances in Microbial Physiology*, volume 52, pages 229–296. Elsevier, 2006.

[27] R. A. Hut and D. G. M. Beersma. Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1574):2141–2154, 2011.

[28] Steven A. Brown, Elzbieta Kowalska, and Robert Dallmann. (Re)inventing the Circadian Feedback Loop. *Developmental Cell*, 22(3):477–487, 2012.

[29] Sarah Lück and Pål O. Westermark. Circadian mRNA expression: insights from modeling and transcriptomics. *Cellular and Molecular Life Sciences*, 73:497–521, 2016.

[30] Y. Miyamoto and A. Sancar. Vitamin B2-based blue-light photoreceptors in the retinohypothalamic tract as the photoactive pigments for setting the circadian clock in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11):6097–6102, 1998.

[31] R. J. Thresher, M. H. Vitaterna, Y. Miyamoto, A. Kazantsev, D. S. Hsu, C. Petit, C. P. Selby, L. Dawut, O. Smithies, J. S. Takahashi, and A. Sancar. Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses. *Science*, 282(5393):1490–1494, 1998.

[32] R. Stanewsky, M. Kaneko, P. Emery, B. Beretta, K. Wager-Smith, S. A. Kay, M. Rosbash, and J. C. Hall. The cryb mutation identifies cryptochrome as a circadian photoreceptor in Drosophila. *Cell*, 95(5):681–692, 1998.

[33] P. Emery, W. V. So, M. Kaneko, J. C. Hall, and M. Rosbash. CRY, a Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, 95(5):669–679, 1998.

[34] N. Öztürk, S.-H. Song, S. Özgür, C. P. Selby, L. Morrison, C. Partch, D. Zhong, and A. Sancar. Structure and Function of Animal Cryptochromes. *Cold Spring Harbor Symposia on Quantitative Biology*, 72(1):119–131, 2007.

[35] Rachel S. Edgar, Edward W. Green, Yuwei Zhao, Gerben van Ooijen, Maria Olmedo, Ximing Qin, Yao Xu, Min Pan, Utham K. Valekunja, Kevin A. Feeney, Elizabeth S. Maywood, Michael H. Hastings, Nitin S. Baliga, Martha Merrow, Andrew J. Millar, Carl H. Johnson, Charalambos P. Kyriacou, John S. O'Neill, and Akhilesh B. Reddy. Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, 2012.

[36] Walter Gehring and Michael Rosbash. The Coevolution of Blue-Light Photoreception and Circadian Rhythms. *Journal of Molecular Evolution*, 57(0):S286–S289, 2003.

[37] E. Tauber. Clock Gene Evolution and Functional Divergence. *Journal of Biological Rhythms*, 19(5):445–458, 2004.

[38] Kaspar H. von Meyenburg. Energetics of the budding cycle of Saccharomyces cerevisiae during glucose limited aerobic growth. *Archiv für Mikrobiologie*, 66(4):289–303, 1969.

[39] R. R. Klevecz, J. Bolen, G. Forrest, and D. B. Murray. A genomewide oscillation in transcription gates DNA replication and cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5):1200–1205, 2004.

[40] N. Slavov and D. Botstein. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular Biology of the Cell*, 22(12):1997–2009, 2011.

[41] Alexandros Papagiannakis, Bastian Niebel, Ernst C. Wit, and Matthias Heinemann. Autonomous Metabolic Oscillations Robustly Gate the Early and Late Cell Cycle. *Molecular Cell*, 65(2):285–295, 2016.

[42] Helen C. Causton, Kevin A. Feeney, Christine A. Ziegler, and John S. O'Neill. Metabolic Cycles in Yeast Share Features Conserved among Circadian Rhythms. *Current biology: CB*, 25(8):1056–1062, 2015.

[43] David Lloyd and Douglas B. Murray. Redox rhythmicity: clocks at the core of temporal coherence. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 29(5):465–473, 2007.

[44] R. R. Klevecz and C. M. Li. Evolution of the clock from yeast to man by period-doubling folds in the cellular oscillator. *Cold Spring Harbor Symposia on Quantitative Biology*, 72:421–429, 2007.

[45] R. K. Finn and R. E. Wilson. Fermentation Process Control, Population Dynamics of a Continuous Propagator for Microorganisms. *Journal of Agricultural and Food Chemistry*, 2:66–69, 1954.

[46] Nikolai Slavov, Joanna Macinskas, Amy Caudy, and David Botstein. Metabolic cycling without cell division cycling in respiring yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 108(47):19090–19095, 2011.

[47] A. J. Burnetti, M. Aydin, and N. E. Buchler. Cell Cycle Start is coupled to Entry into the Yeast Metabolic Cycle Across Diverse Strains and Growth Rates. *Molecular Biology of the Cell*, 27(1):64–74, 2016.

[48] Martin T. Küenzi and Armin Fiechter. Changes in carbohydrate composition and trehalase-activity during the budding cycle of Saccharomyces cerevisiae. *Archiv für Mikrobiologie*, 64(4):396–407, 1969.

[49] Rainer Machné and Douglas B. Murray. The Yin and Yang of Yeast Transcription: Elements of a Global Feedback System between Metabolism and Chromatin. *PLoS ONE*, 7(6):e37906, 2012.

[50] Eric E. Zhang, Andrew C. Liu, Tsuyoshi Hirota, Loren J. Miraglia, Genevieve Welch, Pagkapol Y. Pongsawakul, Xianzhong Liu, Ann Atwood, Jon W. Huss, Jeff Janes, Andrew I. Su, John B. Hogenesch, and Steve A. Kay. A Genome-wide RNAi Screen for Modifiers of the Circadian Clock in Human Cells. *Cell*, 139(1):199–210, 2009.

[51] P. E. Hardin, J. C. Hall, and M. Rosbash. Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels. *Nature*, 343(6258):536–540, 1990.

[52] B. D. Aronson, K. A. Johnson, J. J. Loros, and J. C. Dunlap. Negative feedback defining a circadian clock: autoregulation of the clock gene frequency. *Science*, 263(5153):1578–1584, 1994.

[53] Polly Yingshan Hsu and Stacey L. Harmer. Wheels within wheels: the plant circadian system. *Trends in Plant Science*, 19(4):240–249, 2014.

[54] J. S. Griffith. Mathematics of cellular control processes. I. Negative feedback to one gene. *Journal of Theoretical Biology*, 20(2):202–208, 1968.

[55] P. Rapp. Analysis of biochemical phase shift oscillators by a harmonic balancing technique. *Journal of Mathematical Biology*, 3(3-4):203–224, 1976.

[56] Béla Novák and John J. Tyson. Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology*, 9(12):981–991, 2008.

[57] B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–438, 1965.

[58] Stefan Müller, Josef Hofbauer, Lukas Endler, Christoph Flamm, Stefanie Widder, and Peter Schuster. A generalized model of the repressilator. *Journal of Mathematical Biology*, 53(6):905–937, 2006.

[59] Bharath Ananthasubramaniam and Hanspeter Herzel. Positive Feedback Promotes Oscillations in Negative Feedback Loops. *PLoS ONE*, 9(8):e104761, 2014.

[60] Phillip L. Lowrey and Joseph S. Takahashi. MAMMALIAN CIRCADIAN BIOLOGY: Elucidating Genome-Wide Levels of Temporal Organization. *Annual Review of Genomics and Human Genetics*, 5(1):407–441, 2004.

[61] M. K. Bunger, L. D. Wilsbacher, S. M. Moran, C. Clendenin, L. A. Radcliffe, J. B. Hogenesch, M. C. Simon, J. S. Takahashi, and C. A. Bradfield. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell*, 103(7):1009–1017, 2000.

[62] Jason P DeBruyne, David R Weaver, and Steven M Reppert. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nature Neuroscience*, 10(5):543–545, 2007.

[63] K. Bae, X. Jin, E. S. Maywood, M. H. Hastings, S. M. Reppert, and D. R. Weaver. Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. *Neuron*, 30(2):525–536, 2001.

[64] B. Zheng, U. Albrecht, K. Kaasik, M. Sage, W. Lu, S. Vaishnav, Q. Li, Z. S. Sun, G. Eichele, A. Bradley, and C. C. Lee. Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock. *Cell*, 105(5):683–694, 2001.

[65] G. T. van der Horst, M. Muijtjens, K. Kobayashi, R. Takano, S. Kanno, M. Takao, J. de Wit, A. Verkerk, A. P. Eker, D. van Leenen, R. Buijs, D. Bootsma, J. H. Hoeijmakers, and A. Yasui. Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature*, 398(6728):627–630, 1999.

[66] M. H. Vitaterna, C. P. Selby, T. Todo, H. Niwa, C. Thompson, E. M. Fruechte, K. Hitomi, R. J. Thresher, T. Ishikawa, J. Miyazaki, J. S. Takahashi, and A. Sancar. Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proceedings of the National Academy of Sciences of the United States of America*, 96(21):12114–12119, 1999.

[67] J. Patrick Pett, Anja Korenčič, Felix Wesener, Achim Kramer, and Hanspeter Herzel. Feedback Loops of the Mammalian Circadian Clock Constitute Repressilator. *PLOS Computational Biology*, 12(12):e1005266, 2016.

[68] Anja Korenčič, Grigory Bordyugov, Rok Košir, Damjana Rozman, Marko Goličnik, and Hanspeter Herzel. The Interplay of cis-Regulatory Elements Rules Circadian Rhythms in Mouse Liver. *PLoS ONE*, 7(11):e46835, 2012.

[69] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[70] N. Koike, S.-H. Yoo, H.-C. Huang, V. Kumar, C. Lee, T.-K. Kim, and J. S. Takahashi. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science*, 338(6105):349–354, 2012.

[71] John S. O'Neill, Gerben van Ooijen, Laura E. Dixon, Carl Troein, Florence Corellou, François-Yves Bouget, Akhilesh B. Reddy, and Andrew J. Millar. Circadian rhythms persist without transcription in a eukaryote. *Nature*, 469(7331):554–558, 2011.

[72] Bora Zivkovic. Circadian clock without DNA–History and the power of metaphor. Accessed: 2017-01-09.

[73] Michael Menaker, Zachary C Murphy, and Michael T Sellix. Central control of peripheral circadian oscillators. *Current Opinion in Neurobiology*, 23(5):741–746, 2013.

[74] Jennifer A. Mohawk, Carla B. Green, and Joseph S. Takahashi. Central an peripheral clocks in mammals. *Annual review of neuroscience*, 35:445–462, 2012.

[75] Erik D. Herzog, Sara J. Aton, Rika Numano, Yoshiyuki Sakaki, and Hajime Tei. Temporal Precision in the Mammalian Circadian System: A Reliable Clock from Less Reliable Neurons. *Journal of Biological Rhythms*, 19(1):35–46, 2004.

[76] Charna Dibner, Ueli Schibler, and Urs Albrecht. The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. *Annual Review of Physiology*, 72(1):517–549, 2010.

[77] Natsuko Inagaki, Sato Honma, Daisuke Ono, Yusuke Tanahashi, and Ken-ichi Honma. Separate oscillating cell groups in mouse suprachiasmatic nucleus couple photoperiodically to the onset and end of daily activity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7664–7669, 2007.

[78] Henk Tjebbe VanderLeest, Thijs Houben, Stephan Michel, Tom Deboer, Henk Albus, Mariska J. Vansteensel, Gene D. Block, and Johanna H. Meijer. Seasonal encoding by the circadian pacemaker of the SCN. *Current Biology*, 17(5):468–473, 2007.

[79] Christoph Schmal, Jihwan Myung, Hanspeter Herzel, and Grigory Bordyugov. A Theoretical Study on Seasonality. *Frontiers in Neurology*, 6, 2015.

[80] F. K. Stephan and I. Zucker. Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. *Proceedings of the National Academy of Sciences of the United States of America*, 69(6):1583–1586, 1972.

[81] M. N. Lehman, R. Silver, W. R. Gladstone, R. M. Kahn, M. Gibson, and E. L. Bittman. Circadian rhythmicity restored by neural transplant. Immunocytochemical characterization of the graft and its integration with the host brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 7(6):1626–1638, 1987.

[82] M. R. Ralph, R. G. Foster, F. C. Davis, and M. Menaker. Transplanted suprachiasmatic nucleus determines circadian period. *Science*, 247(4945):975–978, 1990.

[83] F. K. Stephan, J. M. Swann, and C. L. Sisk. Entrainment of circadian rhythms by feeding schedules in rats with suprachiasmatic lesions. *Behavioral and Neural Biology*, 25(4):545–554, 1979.

[84] D. T. Krieger, H. Hauser, and L. C. Krey. Suprachiasmatic nuclear lesions do not abolish food-shifted circadian adrenal and temperature rhythmicity. *Science*, 197(4301):398–399, 1977.

[85] Ozgür Tataroglu, Alec J. Davidson, Luke J. Benvenuto, and Michael Menaker. The methamphetamine-sensitive circadian oscillator (MASCO) in mice. *Journal of Biological Rhythms*, 21(3):185–194, 2006.

[86] Pinar Pezuk, Jennifer A. Mohawk, Tomoko Yoshikawa, Michael T. Sellix, and Michael Menaker. Circadian organization is governed by extra-SCN pacemakers. *Journal of Biological Rhythms*, 25(6):432–441, 2010.

[87] K. Honma, S. Honma, and T. Hiroshige. Activity rhythms in the circadian domain appear in suprachiasmatic nuclei lesioned rats given methamphetamine. *Physiology & Behavior*, 40(6):767–774, 1987.

[88] Francesca Damiola, Nguyet Le Minh, Nicolas Preitner, Benoît Kornmann, Fabienne Fleury-Olela, and Ueli Schibler. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes & Development*, 14(23):2950–2961, 2000.

[89] Jennifer A. Mohawk, Matthew L. Baer, and Michael Menaker. The methamphetamine-sensitive circadian oscillator does not employ canonical clock genes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3519–3524, 2009.

[90] Julie S. Pendergast, Wataru Nakamura, Rio C. Friday, Fumiyuki Hatanaka, Toru Takumi, and Shin Yamazaki. Robust Food Anticipatory Activity in BMAL1-Deficient Mice. *PLoS ONE*, 4(3):e4860, 2009.

[91] Joseph Bass and Joseph S. Takahashi. Circadian Integration of Metabolism and Energetics. *Science*, 330(6009):1349–1354, 2010.

[92] Jacob Richards and Michelle L. Gumz. Advances in understanding the peripheral circadian clocks. *The FASEB Journal*, 26(9):3602–3613, 2012.

[93] Satchidananda Panda. Circadian physiology of metabolism. *Science*, 354(6315):1008–1015, 2016.

[94] Ray Zhang, Nicholas F. Lahens, Heather I. Ballance, Michael E. Hughes, and John B. Hogenesch. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45):16219–16224, 2014.

[95] S. Yamazaki, R. Numano, M. Abe, A. Hida, R. Takahashi, M. Ueda, G. D. Block, Y. Sakaki, M. Menaker, and H. Tei. Resetting central and peripheral circadian oscillators in transgenic rats. *Science*, 288(5466):682–685, 2000.

[96] David K. Welsh, Seung-Hee Yoo, Andrew C. Liu, Joseph S. Takahashi, and Steve A. Kay. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Current Biology*, 14(24):2289–2295, 2004.

[97] Emi Nagoshi, Camille Saini, Christoph Bauer, Thierry Laroche, Felix Naef, and Ueli Schibler. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell*, 119(5):693–705, 2004.

[98] Laura Lande-Diner, Jacob Stewart-Ornstein, Charles J. Weitz, and Galit Lahav. Single-cell analysis of circadian dynamics in tissue explants. *Molecular Biology of the Cell*, 26(22):3940–3945, 2015.

[99] Pål O. Westermark, David K. Welsh, Hitoshi Okamura, and Hanspeter Herzel. Quantification of Circadian Rhythms in Single Cells. *PLoS Computational Biology*, 5(11), 2009.

[100] Ruud M. Buijs, Joke Wortel, Joop J. Van Heerikhuize, Matthijs G. P. Feenstra, Gert J. Ter Horst, Herms J. Romijn, and Andries Kalsbeek. Anatomical and functional demonstration of a multisynaptic suprachiasmatic nucleus adrenal (cortex) pathway: Suprachiasmatic adrenal interaction. *European Journal of Neuroscience*, 11(5):1535–1544, 1999.

[101] A. Kalsbeek, S. Perreau-Lenz, and R. M. Buijs. A Network of (Autonomic) Clock Outputs. *Chronobiology International*, 23(3), 2006.

[102] Cathy Cailotto, Jun Lei, Jan van der Vliet, Caroline van Heijningen, Corbert G. van Eden, Andries Kalsbeek, Paul Pévet, and Ruud M. Buijs. Effects of Nocturnal Light on (Clock) Gene Expression in Peripheral Organs: A Role for the Autonomic Innervation of the Liver. *PLoS ONE*, 4(5):e5650, 2009.

[103] Inés Pineda Torra, Vladimir Tsibulsky, Franck Delaunay, Régis Saladin, Vincent Laudet, Jean-Charles Fruchart, Vladimir Kosykh, and Bart Staels. Circadian and Glucocorticoid Regulation of Rev-erbα Expression in Liver [1]. *Endocrinology*, 141(10):3799–3806, 2000.

[104] Mitsugu Sujino, Keiichi Furukawa, Satoshi Koinuma, Atsuko Fujioka, Mamoru Nagano, Masayuki Iigo, and Yasufumi Shigeyoshi. Differential Entrainment of Peripheral Clocks in the Rat by Glucocorticoid and Feeding. *Endocrinology*, 153(5):2277–2286, 2012.

[105] Pinar Pezük, Jennifer A. Mohawk, Laura A. Wang, and Michael Menaker. Glucocorticoids as Entraining Signals for Peripheral Circadian Oscillators. *Endocrinology*, 153(10):4775–4783, 2012.

[106] Michael E. Hughes, Luciano DiTacchio, Kevin R. Hayes, Christopher Vollmers, S. Pulivarthy, Julie E. Baggs, Satchidananda Panda, and John B. Hogenesch. Harmonics of Circadian Gene Transcription in Mammals. *PLoS Genetics*, 5(4):e1000442, 2009.

[107] Clara Bien Peek, Daniel C. Levine, Jonathan Cedernaes, Akihiko Taguchi, Yumiko Kobayashi, Stacy J. Tsai, Nicolle A. Bonar, Maureen R. McNulty, Kathryn Moynihan Ramsey, and Joseph Bass. Circadian Clock Interaction with HIF1α Mediates Oxygenic Metabolism and Anaerobic Glycolysis in Skeletal Muscle. *Cell Metabolism*, 25(1):86–92, 2016.

[108] R. Daniel Rudic, Peter McNamara, Anne-Maria Curtis, Raymond C. Boston, Satchidananda Panda, John B. Hogenesch, and Garret A. Fitzgerald. BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS biology*, 2(11):e377, 2004.

[109] Katja A. Lamia, Kai-Florian Storch, and Charles J. Weitz. Physiological significance of a peripheral tissue circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39):15172–15177, 2008.

[110] Matthew M. Molusky, Siming Li, Di Ma, Lei Yu, and Jiandie D. Lin. Ubiquitin-Specific Protease 2 Regulates Hepatic Gluconeogenesis and Diurnal Glucose Metabolism Through 11β-Hydroxysteroid Dehydrogenase 1. *Diabetes*, 61(5):1025–1035, 2012.

[111] C. B. Peek, A. H. Affinati, K. M. Ramsey, H.-Y. Kuo, W. Yu, L. A. Sena, O. Ilkayeva, B. Marcheva, Y. Kobayashi, C. Omura, D. C. Levine, D. J. Bacsik, D. Gius, C. B. Newgard, E. Goetzman, N. S. Chandel, J. M. Denu, M. Mrksich, and J. Bass. Circadian Clock NAD+ Cycle Drives Mitochondrial Oxidative Metabolism in Mice. *Science*, 342(6158):1243417–1243417, 2013.

[112] Joshua J Gooley. Circadian regulation of lipid metabolism. *Proceedings of the Nutrition Society*, 75(04):440–450, 2016.

[113] David Jacobi, Sihao Liu, Kristopher Burkewitz, Nora Kory, Nelson H. Knudsen, Ryan K. Alexander, Ugur Unluturk, Xiaobo Li, Xiaohui Kong, Alexander L. Hyde, Matthew R. Gangl, William B. Mair, and Chih-Hao Lee. Hepatic Bmal1 Regulates Rhythmic Mitochondrial Dynamics and Promotes Metabolic Fitness. *Cell Metabolism*, 22(4), 2015.

[114] Olga Cela, Rosella Scrima, Valerio Pazienza, Giuseppe Merla, Giorgia Benegiamo, Bartolomeo Augello, Sabino Fugetto, Marta Menga, Rosa Rubino, Luise Fuhr, Angela Relogio, Claudia Piccoli, Gianluigi Mazzoccoli, and Nazzareno Capitanio. Clock genes-dependent acetylation of complex I sets rhythmic activity of mitochondrial OxPhos. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1863(4):596–606, 2016.

[115] Adi Neufeld-Cohen, Maria S. Robles, Rona Aviram, Gal Manella, Yaarit Adamovich, Benjamin Ladeuix, Dana Nir, Liat Rousso-Noori, Yael Kuperman, Marina Golik, Matthias Mann, and Gad Asher. Circadian control of oscillations in mitochondrial rate-limiting enzymes and nutrient utilization by PERIOD proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12):E1673–E1682, 2016.

[116] Benedikt Westermann. Bioenergetic role of mitochondrial fusion and fission. *Biochimica et Biophysica Acta - Bioenergetics*, 1817(10):1833–1838, 2012.

[117] S. Yamazaki, Y. Ishida, and S. Inouye. Circadian rhythms of adenosine triphosphate contents in the suprachiasmatic nucleus, anterior hypothalamic area and caudate putamen of the rat–negative correlation with electrical activity. *Brain Research*, 664(1-2):237–240, 1994.

[118] Steven A. Brown. Circadian Metabolism: From Mechanisms to Metabolomics and Medicine. *Trends in Endocrinology & Metabolism*, 27(6):415–426, 2016.

[119] Uma M. Sachdeva and Craig B. Thompson. Diurnal rhythms of autophagy: implications for cell biology and human disease. *Autophagy*, 4(5):581–589, 2008.

[120] Marrit Putker and John Stuart O'Neill. Reciprocal Control of the Circadian Clock and Cellular Redox State - a Critical Appraisal. *Molecules and Cells*, 39(1):6–19, 2016.

[121] Adam R. Wende, Martin E. Young, John Chatham, Jianhua Zhang, Namakkal S. Rajasekaran, and Victor M. Darley-Usmar. Redox biology and the interface between bioenergetics, autophagy and circadian control of metabolism. *Free Radical Biology and Medicine*, 100:94–107, 2016.

[122] Di Ma, Satchidananda Panda, and Jiandie D Lin. Temporal orchestration of circadian autophagy rhythm by C/EBP[beta]. *EMBO Journal*, 30:4642–4651, 2011.

[123] Selma Masri, Marlene Cervantes, and Paolo Sassone-Corsi. The circadian clock and cell cycle: interconnected biological circuits. *Current Opinion in Cell Biology*, 25(6):730–734, 2013.

[124] J. Bieler, R. Cannavo, K. Gustafson, C. Gobet, D. Gatfield, and F. Naef. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular Systems Biology*, 10(7):739–739, 2014.

[125] K-F Storch, O Lipan, I Leykin, N Viswanathan, F C Davis, W H Wong, and C J Weitz. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83, 2002.

[126] S Panda, M P Antoch, B H Miller, A I Su, A B Schook, M Straume, P G Schultz, S A Kay, J S Takahashi, and J B Hogenesch. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, 109(3):307–320, 2002.

[127] Brooke H. Miller, Erin L. McDearmon, Satchidananda Panda, Kevin R. Hayes, Jie Zhang, Jessica L. Andrews, Marina P. Antoch, John R. Walker, Karyn A. Esser, John B. Hogenesch, and Joseph S. Takahashi. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9):3342–3347, 2007.

[128] Gwendal Le Martelot, Thierry Claudel, David Gatfield, Olivier Schaad, Benoît Kornmann, Giuseppe Lo Sasso, Antonio Moschetta, and Ueli Schibler. REV-ERBalpha participates in circadian SREBP signaling and bile acid homeostasis. *PLoS biology*, 7(9):e1000181, 2009.

[129] Christopher Vollmers, Shubhroz Gill, Luciano DiTacchio, Sandhya R. Pulivarthy, Hiep D. Le, and Satchidananda Panda. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50):21453–21458, 2009.

[130] K. L. Eckel-Mahan, V. R. Patel, R. P. Mohney, K. S. Vignola, P. Baldi, and P. Sassone-Corsi. Coordination of the transcriptome and metabolome by the circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, 109(14):5541–5546, 2012.

[131] Han Cho, Xuan Zhao, Megumi Hatori, Ruth T. Yu, Grant D. Barish, Michael T. Lam, Ling-Wa Chong, Luciano DiTacchio, Annette R. Atkins, Christopher K. Glass, Christopher Liddle, Johan Auwerx, Michael Downes, Satchidananda Panda, and Ronald M. Evans. Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β. *Nature*, 485(7396):123–127, 2012.

[132] Selma Masri, Vishal R. Patel, Kristin L. Eckel-Mahan, Shahaf Peleg, Ignasi Forne, Andreas G. Ladurner, Pierre Baldi, Axel Imhof, and Paolo Sassone-Corsi. Circadian acetylome reveals regulation of mitochondrial metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 110(9):3339–3344, 2013.

[133] Kristin L. Eckel-Mahan, Vishal R. Patel, Sara de Mateo, Ricardo Orozco-Solis, Nicholas J. Ceglia, Saurabh Sahar, Sherry A. Dilag-Penilla, Kenneth A. Dyar, Pierre Baldi, and Paolo Sassone-Corsi. Reprogramming of the Circadian Clock by Nutritional Challenge. *Cell*, 155(7):1464–1478, 2013.

[134] Akhilesh B. Reddy, Natasha A. Karp, Elizabeth S. Maywood, Elizabeth A. Sage, Michael Deery, John S. O'Neill, Gabriel K.Y. Wong, Jo Chesham, Mark Odell, Kathryn S. Lilley, Charalambos P. Kyriacou, and Michael H. Hastings. Circadian Orchestration of the Hepatic Proteome. *Current Biology*, 16(11):1107–1115, 2006.

[135] D. Mauvoisin, J. Wang, C. Jouffe, E. Martin, F. Atger, P. Waridel, M. Quadroni, F. Gachon, and F. Naef. Circadian clock-dependent and -independent rhythmic proteomes

implement distinct diurnal functions in mouse liver. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1):167–172, 2014.

[136] Maria S. Robles, Jürgen Cox, and Matthias Mann. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLoS Genetics*, 10(1):e1004047, 2014.

[137] Jingkui Wang, Daniel Mauvoisin, Eva Martin, Florian Atger, Antonio Nùñes Galindo, Loïc Dayon, Federico Sizzano, Alessio Palini, Martin Kussmann, Patrice Waridel, Manfredo Quadroni, Vjekoslav Dulić, Felix Naef, and Frédéric Gachon. Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. *Cell Metabolism*, 2016.

[138] Megumi Hatori, Christopher Vollmers, Amir Zarrinpar, Luciano DiTacchio, Eric A. Bushong, Shubhroz Gill, Mathias Leblanc, Amandine Chaix, Matthew Joens, James A.J. Fitzpatrick, Mark H. Ellisman, and Satchidananda Panda. Time-Restricted Feeding without Reducing Caloric Intake Prevents Metabolic Diseases in Mice Fed a High-Fat Diet. *Cell Metabolism*, 15(6):848–860, 2012.

[139] Jean-Michel Fustin, Masao Doi, Hiroyuki Yamada, Rie Komatsu, Shigeki Shimba, and Hitoshi Okamura. Rhythmic Nucleotide Synthesis in the Liver: Temporal Segregation of Metabolites. *Cell Reports*, 1(4):341–349, 2012.

[140] Koichi Fujisawa, Taro Takami, Yoshitaka Kimoto, Toshihiko Matsumoto, Naoki Yamamoto, Shuji Terai, and Isao Sakaida. Circadian variations in the liver metabolites of medaka (Oryzias latipes). *Scientific Reports*, 6:20916, 2016.

[141] Melanie Tran, Zhihong Yang, Suthat Liangpunsakul, and Li Wang. Metabolomics Analysis Revealed Distinct Cyclic Changes of Metabolites Altered by Chronic Ethanol-Plus-Binge and *Shp* Deficiency. *Alcoholism: Clinical and Experimental Research*, 40(12):2548–2556, 2016.

[142] Hans Reinke and Gad Asher. Circadian Clock Control of Liver Metabolic Functions. *Gastroenterology*, 150(3):574–580, 2016.

[143] M. G. Koopman, G. C. Koomen, R. T. Krediet, E. A. de Moor, F. J. Hoek, and L. Arisz. Circadian rhythm of glomerular filtration rate in normal individuals. *Clinical Science (London, England: 1979)*, 77(1):105–111, 1989.

[144] M. Pons, O. Forpomès, S. Espagnet, and J. Cambar. Relationship between circadian changes in renal hemodynamics and circadian changes in urinary glycosaminoglycan excretion in normal rats. *Chronobiology International*, 13(5):349–358, 1996.

[145] L. R. Stow and M. L. Gumz. The Circadian Clock in the Kidney. *Journal of the American Society of Nephrology*, 22(4):598–604, 2011.

[146] Mohammad Saifur Rohman, Noriaki Emoto, Hidemi Nonaka, Ryusuke Okura, Masataka Nishimura, Kazuhiro Yagita, Gijsbertus T.J. Van Der Horst, Masafumi Matsuo, Hitoshi

Okamura, and Mitsuhiro Yokoyama. Circadian clock genes directly regulate expression of the Na+/H+ exchanger NHE3 in the kidney. *Kidney International*, 67(4):1410–1419, 2005.

[147] A. M. Zuber, G. Centeno, S. Pradervand, S. Nikolaeva, L. Maquelin, L. Cardinaux, O. Bonny, and D. Firsov. Molecular clock is involved in predictive circadian adjustment of renal function. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16523–16528, 2009.

[148] Michelle L. Gumz, Lisa R. Stow, I. Jeanette Lynch, Megan M. Greenlee, Alicia Rudin, Brian D. Cain, David R. Weaver, and Charles S. Wingo. The circadian clock protein Period 1 regulates expression of the renal epithelial sodium channel in mice. *Journal of Clinical Investigation*, 119(8):2423–2434, 2009.

[149] K. Solocinski and M. L. Gumz. The Circadian Clock in the Regulation of Renal Rhythms. *Journal of Biological Rhythms*, 30(6):470–486, 2015.

[150] Melissa J. Moore. From birth to death: the complex lives of eukaryotic mRNAs. *Science*, 309(5740):1514–1518, 2005.

[151] Stefanie Gerstberger, Markus Hafner, and Thomas Tuschl. A census of human RNA-binding proteins. *Nature Reviews Genetics*, 15(12):829–845, 2014.

[152] Eric Londin, Phillipe Loher, Aristeidis G. Telonis, Kevin Quann, Peter Clark, Yi Jing, Eleftheria Hatzimichael, Yohei Kirino, Shozo Honda, Michelle Lally, Bharat Ramratnam, Clay E. S. Comstock, Karen E. Knudsen, Leonard Gomella, George L. Spaeth, Lisa Hark, L. Jay Katz, Agnieszka Witkiewicz, Abdolmohamad Rostami, Sergio A. Jimenez, Michael A. Hollingsworth, Jen Jen Yeh, Chad A. Shaw, Steven E. McKenzie, Paul Bray, Peter T. Nelson, Simona Zupo, Katrien Van Roosbroeck, Michael J. Keating, George A. Calin, Charles Yeo, Masaya Jimbo, Joseph Cozzitorto, Jonathan R. Brody, Kathleen Delgrosso, John S. Mattick, Paolo Fortina, and Isidore Rigoutsos. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10):E1106–E1115, 2015.

[153] Jacek Krol, Inga Loedige, and Witold Filipowicz. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics*, 11(9):597–610, 2010.

[154] Julia Berretta and Antonin Morillon. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Reports*, 10(9):973–982, 2009.

[155] B. G. Robinson, D. M. Frim, W. J. Schwartz, and J. A. Majzoub. Vasopressin mRNA in the suprachiasmatic nuclei: daily regulation of polyadenylate tail length. *Science*, 241(4863):342–344, 1988.

[156] W.V. So and M. Rosbash. Post-transcriptional regulation contributes to Drosophila clock gene mRNA cycling. *EMBO Journal*, 16(23):7146–7155, 1997.

[157] Y. Cheng, B. Gvakharia, and P. E. Hardin. Two alternatively spliced transcripts from the Drosophila period gene rescue rhythms having different molecular and behavioral characteristics. *Molecular and Cellular Biology*, 18(11):6505–6514, 1998.

[158] John Majercak, David Sidote, Paul E. Hardin, and Isaac Edery. How a Circadian Clock Adapts to Seasonal Decreases in Temperature and Day Length. *Neuron*, 24(1):219–230, 1999.

[159] E. Kwak. Essential Role of 3'-Untranslated Region-mediated mRNA Decay in Circadian Oscillations of Mouse Period3 mRNA. *Journal of Biological Chemistry*, 281(28):19100–19106, 2006.

[160] K.-C. Woo, T.-D. Kim, K.-H. Lee, D.-Y. Kim, W. Kim, K.-Y. Lee, and K.-T. Kim. Mouse period 2 mRNA circadian oscillation is modulated by PTB-mediated rhythmic mRNA degradation. *Nucleic Acids Research*, 37(1):26–37, 2009.

[161] K.-C. Woo, D.-C. Ha, K.-H. Lee, D.-Y. Kim, T.-D. Kim, and K.-T. Kim. Circadian Amplitude of Cryptochrome 1 Is Modulated by mRNA Stability Regulation via Cytoplasmic hnRNP D Oscillation. *Molecular and Cellular Biology*, 30(1):197–205, 2010.

[162] S. Kojima, K. Matsumoto, M. Hirose, M. Shimada, M. Nagano, Y. Shigeyoshi, S.-i. Hoshino, K. Ui-Tei, K. Saigo, C. B. Green, Y. Sakaki, and H. Tei. LARK activates posttranscriptional expression of an essential mammalian clock protein, PERIOD1. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1859–1864, 2007.

[163] M. Andrea Markus and Brian J. Morris. RBM4: A multifunctional RNA-binding protein. *The International Journal of Biochemistry & Cell Biology*, 41(4):740–743, 2009.

[164] Carla B. Green and Joseph C. Besharse. Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25):14884–14888, 1996.

[165] Y. Wang, D. L. Osterbur, P. L. Megaw, G. Tosini, C. Fukuhara, C. B. Green, and J. C. Besharse. Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse. *BMC Developmental Biology*, 1:9, 2001.

[166] Alan R. Godwin, Shihoko Kojima, Carla B. Green, and Jeffrey Wilusz. Kiss your tail goodbye: The role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(6-7):571–579, 2013.

[167] Jeremy J. Stubblefield, Jérémy Terrien, and Carla B. Green. Nocturnin: at the crossroads of clocks and metabolism. *Trends in Endocrinology and Metabolism*, 23(7):326–333, 2012.

[168] J. Morf, G. Rey, K. Schneider, M. Stratmann, J. Fujita, F. Naef, and U. Schibler. Cold-Inducible RNA-Binding Protein Modulates Circadian Gene Expression Posttranscriptionally. *Science*, 338(6105):379–383, 2012.

[169] Yuting Liu, Wenchao Hu, Yasuhiro Murakawa, Jingwen Yin, Gang Wang, Markus Landthaler, and Jun Yan. Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Scientific Reports*, 3, 2013.

[170] Ivana Gotic, Saeed Omidi, Fabienne Fleury-Olela, Nacho Molina, Felix Naef, and Ueli Schibler. Temperature regulates splicing efficiency of the cold-inducible RNA-binding protein gene *Cirbp*. *Genes & Development*, 30(17):2005–2017, 2016.

[171] D. Gatfield, G. Le Martelot, C. E. Vejnar, D. Gerlach, O. Schaad, F. Fleury-Olela, A.-L. Ruskeepaa, M. Oresic, C. C. Esau, E. M. Zdobnov, and U. Schibler. Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes & Development*, 23(11):1313–1326, 2009.

[172] Mariana Lagos-Quintana, Reinhard Rauhut, Abdullah Yalcin, Jutta Meyer, Winfried Lendeckel, and Thomas Tuschl. Identification of Tissue-Specific MicroRNAs from Mouse. *Current Biology*, 12(9):735–739, 2002.

[173] Jan Krützfeldt, Nikolaus Rajewsky, Ravi Braich, Kallanthottathil G. Rajeev, Thomas Tuschl, Muthiah Manoharan, and Markus Stoffel. Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, 438(7068):685–689, 2005.

[174] Christine Esau, Scott Davis, Susan F. Murray, Xing Xian Yu, Sanjay K. Pandey, Michael Pear, Lynnetta Watts, Sheri L. Booten, Mark Graham, Robert McKay, Amuthakannan Subramaniam, Stephanie Propp, Bridget A. Lollo, Susan Freier, C. Frank Bennett, Sanjay Bhanot, and Brett P. Monia. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metabolism*, 3(2):87–98, 2006.

[175] Haifang Wang, Zenghua Fan, Meng Zhao, Juan Li, Minghua Lu, Wei Liu, Hao Ying, Mofang Liu, and Jun Yan. Oscillating primary transcripts harbor miRNAs with circadian functions. *Scientific Reports*, 6:21598, 2016.

[176] Sadanand Vodala, Stefan Pescatore, Joseph Rodriguez, Marita Buescher, Ya-Wen Chen, Ruifen Weng, Stephen M. Cohen, and Michael Rosbash. The Oscillating miRNA 959-964 Cluster Impacts Drosophila Feeding Time and Other Circadian Outputs. *Cell Metabolism*, 16(5):601–612, 2012.

[177] Ngoc-Hien Du, Alaaddin Bulak Arpat, Mara De Matos, and David Gatfield. MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife*, 3, 2014.

[178] J. Hausser, A. P. Syed, N. Selevsek, E. van Nimwegen, L. Jaskiewicz, R. Aebersold, and M. Zavolan. Timescales and bottlenecks in miRNA-dependent gene regulation. *Molecular Systems Biology*, 9(1):711–711, 2014.

[179] Sarah Lück, Kevin Thurley, Paul F. Thaben, and Pål O. Westermark. Rhythmic Degradation Explains and Unifies Circadian Transcriptome and Proteome Data. *Cell Reports*, 9(2):741–751, 2014.

[180] Manon Torres, Denis Becquet, Marie-Pierre Blanchard, Séverine Guillen, Bénédicte Boyer, Mathias Moreno, Jean-Louis Franc, and Anne-Marie Fran¸sois Bellan. Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. *eLife*, 5, 2016.

[181] Christine M. Clemson, John N. Hutchinson, Sergio A. Sara, Alexander W. Ensminger, Archa H. Fox, Andrew Chess, and Jeanne B. Lawrence. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Molecular Cell*, 33(6):717–726, 2009.

[182] Yasnory TF Sasaki and Tetsuro Hirose. How to build a paraspeckle. *Genome Biology*, 10(7):227, 2009.

[183] Ling-Ling Chen and Gordon G. Carmichael. Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle*, 7(21):3294–3301, 2008.

[184] Eric T. Wang, Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–476, 2008.

[185] Stefan Mockenhaupt and Eugene V. Makeyev. Non-coding functions of alternative pre-mRNA splicing in development. *Seminars in Cell & Developmental Biology*, 47-48:32–39, 2015.

[186] Nicholas J. McGlincy, Amandine Valomon, Johanna E. Chesham, Elizabeth S. Maywood, Michael H. Hastings, and Jernej Ule. Regulation of alternative splicing by the circadian clock and food related cues. *Genome Biology*, 13(6):R54, 2012.

[187] Nicole L. Garneau, Jeffrey Wilusz, and Carol J. Wilusz. The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology*, 8(2):113–126, 2007.

[188] Xiaokan Zhang, Anders Virtanen, and Frida E Kleiman. To polyadenylate or to deadenylate. *Cell Cycle*, 9(22):4437–4449, 2010.

[189] Shihoko Kojima, Elaine L. Sher-Chen, and Carla B. Green. Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes & Development*, 26(24):2724–2736, 2012.

[190] Shihoko Kojima, Kerry L. Gendreau, Elaine L. Sher-Chen, Peng Gao, and Carla B. Green. Changes in poly(A) tail length dynamics from the loss of the circadian deadenylase Nocturnin. *Scientific Reports*, 5:17059, 2015.

[191] Dan Dominissini, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali Salmon-Divon, Lior Ungar, Sivan Osenberg, Karen Cesarkas, Jasmine Jacob-Hirsch, Ninette Amariglio, Martin Kupiec, Rotem Sorek, and Gideon Rechavi. Topology of the human and mouse m6a RNA methylomes revealed by m6a-seq. *Nature*, 485(7397):201–206, 2012.

[192] Jean-Michel Fustin, Masao Doi, Yoshiaki Yamaguchi, Hayashi Hida, Shinichi Nishimura, Minoru Yoshida, Takayuki Isagawa, Masaki Suimye Morioka, Hideaki Kakeya, Ichiro Manabe, and Hitoshi Okamura. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell*, 155(4):793–806, 2013.

[193] Francesco Ferrari, Artyom A Alekseyenko, Peter J Park, and Mitzi I Kuroda. Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nature Structural & Molecular Biology*, 21(2):118–125, 2014.

[194] Dimos Gaidatzis, Lukas Burger, Maria Florescu, and Michael B Stadler. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nature Biotechnology*, 33(7):722–729, 2015.

[195] Peter J. Park. ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews. Genetics*, 10(10):669–680, 2009.

[196] Nikolay V. Rozhkov. Global Run-On Sequencing (GRO-seq) Library Preparation from Drosophila Ovaries. *Methods in Molecular Biology*, 1328:217–230, 2015.

[197] Y. L. Khodor, J. Rodriguez, K. C. Abruzzi, C.-H. A. Tang, M. T. Marr, and M. Rosbash. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes & Development*, 25(23):2502–2512, 2011.

[198] Mattia Brugiolo, Lydia Herzel, and Karla M Neugebauer. Counting on co-transcriptional splicing. *F1000Prime Reports*, 5, 2013.

[199] J. Rodriguez, C.-H. A. Tang, Y. L. Khodor, S. Vodala, J. S. Menet, and M. Rosbash. Nascent-Seq analysis of Drosophila cycling gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):E275–E284, 2013.

[200] P. F. Thaben and P. O. Westermark. Detecting Rhythms in Time Series with RAIN. *Journal of Biological Rhythms*, 29(6):391–400, 2014.

[201] Halberg. Circadian System Phase - An Aspect of Temporal Morphology; Procedures and Illustrative Examples. In *The Cellular Aspects of Biorhythms*, pages 20–48. Mayersbach, H.v., 1965.

[202] Martin Straume. DNA Microarray Time Series Analysis: Automated Statistical Assessment of Circadian Rhythms in Gene Expression Patterning. In *Methods in Enzymology*, volume 383, pages 149–166. Elsevier, 2004.

[203] S. Wichert, K. Fokianos, and K. Strimmer. Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, 20(1):5–20, 2004.

[204] M. E. Hughes, J. B. Hogenesch, and K. Kornacker. JTK_cycle: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms*, 25(5):372–380, 2010.

[205] Herman Wijnen, Felix Naef, and Michael W. Young. Molecular and Statistical Tools for Circadian Transcript Profiling. In *Methods in Enzymology*, volume 393, pages 341–365. Elsevier, 2005.

[206] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, 57(1):289–300, 1995.

[207] L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(5909):1845–1848, 2008.

[208] Keren Bahar Halpern, Sivan Tanami, Shanie Landen, Michal Chapal, Liran Szlak, Anat Hutzler, Anna Nizhberg, and Shalev Itzkovitz. Bursty Gene Expression in the Intact Mammalian Liver. *Molecular Cell*, 58(1):147–156, 2015.

[209] Jarnail Singh and Richard A Padgett. Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, 16(11):1128–1133, 2009.

[210] E. Wahle. Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *The Journal of Biological Chemistry*, 266(5):3131–3139, 1991.

[211] Marlene Oeffinger and Daniel Zenklusen. To the pore and through the pore: A story of mRNA export kinetics. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(6):494–506, 2012.

[212] C. C. Friedel, L. Dolken, Z. Ruzsics, U. H. Koszinowski, and R. Zimmer. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Research*, 37(17):e115–e115, 2009.

[213] Björn Schwanhäusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011.

[214] Shanrong Zhao, Wai-Ping Fung-Leung, Anton Bittner, Karen Ngo, and Xuejun Liu. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE*, 9(1):e78644, 2014.

[215] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J, 3rd ed edition, 2003.

[216] Florian Atger, Cédric Gobet, Julien Marquis, Eva Martin, Jingkui Wang, Benjamin Weger, Grégory Lefebvre, Patrick Descombes, Felix Naef, and Frédéric Gachon. Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. *Proceedings of the National Academy of Sciences of the United States of America*, 112(47):E6579–E6588, 2015.

[217] Ionita Ghiran. Project description: Impact of circadian rhythm in obtaining reference profiles of exRNAs in healthy in healthy individuals. `http://exrna.org/projects/impact-circadian-rhythm-obtaining-reference-profiles-exrnas.-healthy-healthy-individuals/`. Accessed: 2017-01-07.

[218] R. Sandberg, J. R. Neilson, A. Sarma, P. A. Sharp, and C. B. Burge. Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science*, 320(5883):1643–1647, 2008.

[219] T. Kadowaki, M. Hitomi, S. Chen, and A. M. Tartakoff. Nuclear mRNA accumulation causes nucleolar fragmentation in yeast mtr2 mutant. *Molecular Biology of the Cell*, 5(11):1253–1263, 1994.

[220] D. Weil, S. Boutain, A. Audibert, and F. Dautry. Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression. *RNA* , 6(7):962–975, 2000.

[221] Janusz Niedojadło, Konrad Dełeńko, and Katarzyna Niedojadło. Regulation of poly(A) RNA retention in the nucleus as a survival strategy of plants during hypoxia. *RNA Biology*, 13(5):531–543, 2016.

[222] Jack D. Keene. RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics*, 8(7):533–543, 2007.

[223] Da Wei Huang, Brad T Sherman, and Richard A Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57, 2008.

[224] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1–13, 2009.

[225] Li Wang and Rui Yi. 3'UTRs take a long shot in the brain. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 36(1):39–45, 2014.

[226] Xiao Li, Gerald Quon, Howard D. Lipshitz, and Quaid Morris. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA*, 16(6):1096–1107, 2010.

[227] Michael E. Hughes, Hee-Kyung Hong, Jason L. Chong, Alejandra A. Indacochea, Samuel S. Lee, Michael Han, Joseph S. Takahashi, and John B. Hogenesch. Brain-Specific Rescue of Clock Reveals System-Driven Transcriptional Rhythms in Peripheral Tissue. *PLoS Genetics*, 8(7):e1002835, 2012.

[228] Gaspard Cretenet, Mikaël Le Clech, and Frédéric Gachon. Circadian Clock-Coordinated 12 Hr Period Rhythmic Activation of the IRE1α Pathway Controls Lipid Metabolism in Mouse Liver. *Cell Metabolism*, 11(1):47–57, 2010.

[229] Pål O. Westermark and Hanspeter Herzel. Mechanism for 12 Hr Rhythm Generation by the Circadian Clock. *Cell Reports*, 3(4):1228–1238, 2013.

[230] Martin J Hicks, Chin-Rang Yang, Matthew V Kotlajich, and Klemens J Hertel. Linking Splicing to Pol II Transcription Stabilizes Pre-mRNAs and Influences Splicing Patterns. *PLoS Biology*, 4(6):e147, 2006.

[231] Gad Asher, David Gatfield, Markus Stratmann, Hans Reinke, Charna Dibner, Florian Kreppel, Raul Mostoslavsky, Frederick W. Alt, and Ueli Schibler. SIRT1 Regulates Circadian Clock Gene Expression through PER2 Deacetylation. *Cell*, 134(2):317–328, 2008.

[232] K. A. Lamia, U. M. Sachdeva, L. DiTacchio, E. C. Williams, J. G. Alvarez, D. F. Egan, D. S. Vasquez, H. Juguilon, S. Panda, R. J. Shaw, C. B. Thompson, and R. M. Evans. AMPK Regulates the Circadian Clock by Cryptochrome Phosphorylation and Degradation. *Science*, 326(5951):437–440, 2009.

[233] John J. Tyson and Béla Novák. Functional Motifs in Biochemical Reaction Networks. *Annual Review of Physical Chemistry*, 61(1):219–240, 2010.

[234] P. Ruoff, M. Vinsjevik, C. Monnerjahn, and L. Rensing. The Goodwin oscillator: on the importance of degradation reactions in the circadian clock. *Journal of Biological Rhythms*, 14(6):469–479, 1999.

[235] Paul François, Nicolas Despierre, and Eric D. Siggia. Adaptive Temperature Compensation in Circadian Oscillations. *PLoS Computational Biology*, 8(7):e1002585, 2012.

[236] Peter Ruoff and Ludger Rensing. The Temperature-Compensated Goodwin Model Simulates Many Circadian Clock Properties. *Journal of Theoretical Biology*, 179(4):275–285, 1996.

[237] Zhang Cheng, Feng Liu, Xiao-Peng Zhang, and Wei Wang. Reversible Phosphorylation Subserves Robust Circadian Rhythms by Creating a Switch in Inactivating the Positive Element. *Biophysical Journal*, 97(11):2867–2875, 2009.

[238] Treenut Saithong, Kevin J. Painter, and Andrew J. Millar. The Contributions of Interlocking Loops and Extensive Nonlinearity to the Properties of Circadian Clock Models. *PLoS ONE*, 5(11):e13867, 2010.

[239] Angela Relógio, Pål O. Westermark, Thomas Wallach, Katja Schellenberg, Achim Kramer, and Hanspeter Herzel. Tuning the Mammalian Circadian Clock: Robust Synergy of Two Loops. *PLoS Computational Biology*, 7(12):e1002309, 2011.

[240] Aurore Woller, Hélène Duez, Bart Staels, and Marc Lefranc. A Mathematical Model of the Liver Circadian Clock Linking Feeding and Fasting Cycles to Clock Function. *Cell Reports*, 17(4):1087–1097, 2016.

[241] A. G. McKendrick and M. Kesava Pai. XLV.-The Rate of Multiplication of Micro-organisms: A Mathematical Study. *Proceedings of the Royal Society of Edinburgh*, 31:649–653, 1912.

[242] A. G. McKendrick. Applications of Mathematics to Medical Problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98, 1925.

[243] H. von Foerster. Some Remarks on Changing Populations. *In The Kinetics of Cellular Proliferation*, pages 382–407, 1959.

[244] B. S. Berlett and E. R. Stadtman. Protein Oxidation in Aging, Disease, and Oxidative Stress. *Journal of Biological Chemistry*, 272(33):20313–20316, 1997.

[245] Audrey Desvergne, Nicolas Ugarte, Sabrina Radjei, Monique Gareil, Isabelle Petropoulos, and Bertrand Friguet. Circadian modulation of proteasome activity and accumulation of oxidized protein in human embryonic kidney HEK 293 cells and primary dermal fibroblasts. *Free Radical Biology and Medicine*, 94:195–207, 2016.

[246] Tobias Jung, Nicolle Bader, and Tilman Grune. Oxidized proteins: Intracellular distribution and recognition by the proteasome. *Archives of Biochemistry and Biophysics*, 462(2):231–237, 2007.

[247] George Brawerman. The Role of the Poly(A) Sequence in Mammalian Messenger RN. *Critical Reviews in Biochemistry*, 10(1):1–38, 1981.

[248] Aaron C. Goldstrohm and Marvin Wickens. Multifunctional deadenylase complexes diversify mRNA control. *Nature Reviews Molecular Cell Biology*, 9(4):337–344, 2008.

[249] Jong-Eun Park, Hyerim Yi, Yoosik Kim, Hyeshik Chang, and V. Narry Kim. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Molecular Cell*, 62(3):462–471, 2016.

[250] Mandy Jeske, Claudia Temme, and Elmar Wahle. Assaying mRNA Deadenylation In Vitro. In Joanna Rorbach and Agnieszka J. Bobrowicz, editors, *Polyadenylation*, volume 1125, pages 297–311. Humana Press, Totowa, NJ, 2014.

[251] Claudia Temme and Elmar Wahle. Assaying mRNA Deadenylation In Vivo. In Joanna Rorbach and Agnieszka J. Bobrowicz, editors, *Polyadenylation*, volume 1125, pages 313–324. Humana Press, Totowa, NJ, 2014.

[252] Hyeshik Chang, Jaechul Lim, Minju Ha, and V. Narry Kim. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3' End Modifications. *Molecular Cell*, 53(6):1044–1052, 2014.

[253] Alexander O. Subtelny, Stephen W. Eichhorn, Grace R. Chen, Hazel Sive, and David P. Bartel. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, 508(7494):66–71, 2014.

[254] M. Barahona and J. Dattani. Stochastic models of gene transcription with upstream drives: Exact solution and sample path characterisation, 2016.

[255] G. J. Boender, A. A. de Koeijer, and E. A. J. Fischer. Derivation of a Floquet Formalism within a Natural Framework. *Acta Biotheoretica*, 60(3):303–317, 2012.

[256] Il'ja N. Bronštejn, Konstantin A. Semendjaev, Gerhard Musiol, Heiner Mühlig, and I. N. Bronstein, editors. *Taschenbuch der Mathematik*. Deutsch, Frankfurt am Main, 7., vollst. überarb. und erg. aufl edition, 2008.

[257] H. S. Carslaw. *Introduction to the theory of Fourier's series and integrals*. MacMillan and Co., 1921.

[258] F. Halberg, Y. L. Tong, and E. A. Johnson. Circadian System Phase - An Aspect of Temporal Morphology; Procedures and Illustrative Examples. In H. von Mayersbach, editor, *The Cellular Aspects of Biorhythms*, pages 20–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 1967.

[259] Natalie Gilks, Nancy Kedersha, Maranatha Ayodele, Lily Shen, Georg Stoecklin, Laura M. Dember, and Paul Anderson. Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular Biology of the Cell*, 15(12):5383–5398, 2004.

[260] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, 2012.

[261] Michael Lawrence, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin T. Morgan, and Vincent J. Carey. Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, 9(8):e1003118, 2013.

[262] Lioudmila V. Sharova, Alexei A. Sharov, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru S. H. Ko. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA research: an international journal for rapid publication of reports on genes and genomes*, 16(1):45–58, 2009.

[263] K. B. Cook, H. Kazan, K. Zuberi, Q. Morris, and T. R. Hughes. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research*, 39(Database):D301–D308, 2011.

[264] Debashish Ray, Hilal Kazan, Kate B. Cook, Matthew T. Weirauch, Hamed S. Najafabadi, Xiao Li, Serge Gueroussov, Mihai Albu, Hong Zheng, Ally Yang, Hong Na, Manuel Irimia, Leah H. Matzat, Ryan K. Dale, Sarah A. Smith, Christopher A. Yarosh, Seth M. Kelly, Behnam Nabet, Desirea Mecenas, Weimin Li, Rakesh S. Laishram, Mei Qiao, Howard D. Lipshitz, Fabio Piano, Anita H. Corbett, Russ P. Carstens, Brendan J. Frey, Richard A. Anderson, Kristen W. Lynch, Luiz O. F. Penalva, Elissa P. Lei, Andrew G. Fraser, Benjamin J. Blencowe, Quaid D. Morris, and Timothy R. Hughes. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 499(7457):172–177, 2013.

[265] I. Paz, I. Kosti, M. Ares, M. Cline, and Y. Mandel-Gutfreund. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Research*, 42(W1):W361–W367, 2014.

[266] Andrew Yates, Wasiu Akanni, M. Ridwan Amode, Daniel Barrell, Konstantinos Billis, Denise Carvalho-Silva, Carla Cummins, Peter Clapham, Stephen Fitzgerald, Laurent Gil,

Carlos García Girón, Leo Gordon, Thibaut Hourlier, Sarah E. Hunt, Sophie H. Janacek, Nathan Johnson, Thomas Juettemann, Stephen Keenan, Ilias Lavidas, Fergal J. Martin, Thomas Maurel, William McLaren, Daniel N. Murphy, Rishi Nag, Michael Nuhn, Anne Parker, Mateus Patricio, Miguel Pignatelli, Matthew Rahtz, Harpreet Singh Riat, Daniel Sheppard, Kieron Taylor, Anja Thormann, Alessandro Vullo, Steven P. Wilder, Amonida Zadissa, Ewan Birney, Jennifer Harrow, Matthieu Muffato, Emily Perry, Magali Ruffier, Giulietta Spudich, Stephen J. Trevanion, Fiona Cunningham, Bronwen L. Aken, Daniel R. Zerbino, and Paul Flicek. Ensembl 2016. *Nucleic Acids Research*, 44(D1):D710–D716, 2016.

[267] J. O. Henderson, V. Blanc, and N. O. Davidson. Isolation, characterization and developmental regulation of the human apobec-1 complementation factor (ACF) gene. *Biochimica Et Biophysica Acta*, 1522(1):22–30, 2001.

[268] Valerie Blanc, Jeffrey O. Henderson, Elizabeth P. Newberry, Susan Kennedy, Jianyang Luo, and Nicholas O. Davidson. Targeted Deletion of the Murine apobec-1 Complementation Factor (acf) Gene Results in Embryonic Lethality. *Molecular and Cellular Biology*, 25(16):7260–7269, 2005.

[269] Percy Tumbale, Jessica S. Williams, Matthew J. Schellenberg, Thomas A. Kunkel, and R. Scott Williams. Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature*, 506(7486):111–115, 2014.

[270] Beatriz Garcia-Diaz, Emanuele Barca, Andrea Balreira, Luis C. Lopez, Saba Tadesse, Sindhu Krishna, Ali Naini, Caterina Mariotti, Barbara Castellotti, and Catarina M. Quinzii. Lack of aprataxin impairs mitochondrial functions via downregulation of the APE1/NRF1/NRF2 pathway. *Human Molecular Genetics*, 24(16):4516–4529, 2015.

[271] J. F. Welk, A. Charlesworth, G. D. Smith, and A. M. MacNicol. Identification and characterization of the gene encoding human cytoplasmic polyadenylation element binding protein. *Gene*, 263(1-2):113–120, 2001.

[272] R. Mendez and J. D. Richter. Translational control by CPEB: a means to the end. *Nature Reviews. Molecular Cell Biology*, 2(7):521–529, 2001.

[273] Kourosh Salehi-Ashtiani, Andrej Lupłák, Alexander Litovchick, and Jack W. Szosłak. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313(5794):1788–1792, 2006.

[274] Elena Ortiz-Zapater, David Pineda, Neus Martínez-Bosch, Gonzalo Fernández-Miranda, Mar Iglesias, Francesc Alameda, Mireia Moreno, Carolina Eliscovich, Eduardo Eyras, Francisco X. Real, Raúl Méndez, and Pilar Navarro. Key contribution of CPEB4-mediated translational control to cancer progression. *Nature Medicine*, 18(1):83–90, 2011.

[275] Elena Amendola, Remo Sanges, Antonella Galvan, Nina Dathan, Giacomo Manenti, Giuseppe Ferrandino, Francesca Maria Alvino, Tina Di Palma, Marzia Scarfò, Mariastella Zannini, Tommaso A. Dragani, Mario De Felice, and Roberto Di Lauro. A locus on mouse

chromosome 2 is involved in susceptibility to congenital hypothyroidism and contains an essential gene expressed in thyroid. *Endocrinology*, 151(4):1948–1958, 2010.

[276] Debbie-Jane G. Scarlett, Patries M. Herst, and Michael V. Berridge. Multiple proteins with single activities or a single protein with multiple activities: the conundrum of cell surface NADH oxidoreductases. *Biochimica Et Biophysica Acta*, 1708(1):108–119, 2005.

[277] Caroline Vance, Boris Rogelj, Tibor Hortobágyi, Kurt J. De Vos, Agnes Lumi Nishimura, Jemeen Sreedharan, Xun Hu, Bradley Smith, Deborah Ruddy, Paul Wright, Jeban Ganesalingam, Kelly L. Williams, Vineeta Tripathi, Safa Al-Saraj, Ammar Al-Chalabi, P. Nigel Leigh, Ian P. Blair, Garth Nicholson, Jackie de Belleroche, Jean-Marc Gallo, Christopher C. Miller, and Christopher E. Shaw. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science*, 323(5918):1208–1211, 2009.

[278] Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Richa Saxena, Benjamin F. Voight, Valeriya Lyssenko, Noël P. Burtt, Paul I. W. de Bakker, Hong Chen, Jeffrey J. Roix, Sekar Kathiresan, Joel N. Hirschhorn, Mark J. Daly, Thomas E. Hughes, Leif Groop, David Altshuler, Peter Almgren, Jose C. Florez, Joanne Meyer, Kristin Ardlie, Kristina Bengtsson Boström, Bo Isomaa, Guillaume Lettre, Ulf Lindblad, Helen N. Lyon, Olle Melander, Christopher Newton-Cheh, Peter Nilsson, Marju Orho-Melander, Lennart Råstam, Elizabeth K. Speliotes, Marja-Riitta Taskinen, Tiinamaija Tuomi, Candace Guiducci, Anna Berglund, Joyce Carlson, Lauren Gianniny, Rachel Hackett, Liselotte Hall, Johan Holmkvist, Esa Laurila, Marketa Sjögren, Maria Sterner, Aarti Surti, Margareta Svensson, Malin Svensson, Ryan Tewhey, Brendan Blumenstiel, Melissa Parkin, Matthew Defelice, Rachel Barry, Wendy Brodeur, Jody Camarata, Nancy Chia, Mary Fava, John Gibbons, Bob Handsaker, Claire Healy, Kieu Nguyen, Casey Gates, Carrie Sougnez, Diane Gage, Marcia Nizzari, Stacey B. Gabriel, Gung-Wei Chirn, Qicheng Ma, Hemang Parikh, Delwood Richardson, Darrell Ricke, and Shaun Purcell. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829):1331–1336, 2007.

[279] Lisa Traunmüller, Andrea M. Gomez, Thi-Minh Nguyen, and Peter Scheiffele. Control of neuronal synapse specification by a highly dedicated alternative splicing program. *Science*, 352(6288):982–986, 2016.

[280] Thai H. Ho, Nicolas Charlet-B, Michael G. Poulos, Gopal Singh, Maurice S. Swanson, and Thomas A. Cooper. Muscleblind proteins regulate alternative splicing. *The EMBO journal*, 23(15):3103–3112, 2004.

[281] Jean-Christophe Amé, Catherine Spenlehauer, and Gilbert de Murcia. The PARP superfamily. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 26(8):882–893, 2004.

[282] Alexander N. Chkheidze and Stephen A. Liebhaber. A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins. *Molecular and Cellular Biology*, 23(23):8405–8415, 2003.

[283] Jiandie Lin, Pei-Hsuan Wu, Paul T. Tarr, Katrin S. Lindenberg, Julie St-Pierre, Chen-Yu Zhang, Vamsi K. Mootha, Sibylle Jäger, Claudia R. Vianna, Richard M. Reznick, Libin Cui, Monia Manieri, Mi X. Donovan, Zhidan Wu, Marcus P. Cooper, Melina C. Fan, Lindsay M. Rohas, Ann Marie Zavacki, Saverio Cinti, Gerald I. Shulman, Bradford B. Lowell, Dimitri Krainc, and Bruce M. Spiegelman. Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1alpha null mice. *Cell*, 119(1):121–135, 2004.

[284] Archa H. Fox, Yun Wah Lam, Anthony K. L. Leung, Carol E. Lyon, Jens Andersen, Matthias Mann, and Angus I. Lamond. Paraspeckles: a novel nuclear domain. *Current Biology*, 12(1):13–25, 2002.

[285] Kazuhiro Fukumura, Ayako Kato, Yui Jin, Takashi Ideue, Tetsuro Hirose, Naoyuki Kataoka, Toshinobu Fujiwara, Hiroshi Sakamoto, and Kunio Inoue. Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. *Nucleic Acids Research*, 35(16):5303–5311, 2007.

[286] Simona Pedrotti, Roberta Busà, Claudia Compagnucci, and Claudio Sette. The RNA recognition motif protein RBM11 is a novel tissue-specific splicing regulator. *Nucleic Acids Research*, 40(3):1021–1032, 2012.

[287] Puri Fortes, Dasa Longman, Susan McCracken, Joanna Y. Ip, Raymond Poot, Iain W. Mattaj, Javier F. Cáceres, and Benjamin J. Blencowe. Identification and characterization of RED120: a conserved PWI domain protein with links to splicing and 3'-end formation. *FEBS letters*, 581(16):3087–3097, 2007.

[288] Limin Shu, Wensheng Yan, and Xinbin Chen. RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. *Genes & Development*, 20(21):2961–2972, 2006.

[289] Chenchen Wang, Yuanfan Chen, Hongkui Deng, Shaorong Gao, and Lingsong Li. Rbm46 regulates trophectoderm differentiation by stabilizing Cdx2 mRNA in early mouse embryos. *Stem Cells and Development*, 24(7):904–915, 2015.

[290] Benjamin Cieply, Juw Won Park, Angela Nakauka-Ddamba, Thomas W. Bebee, Yang Guo, Xuequn Shang, Christopher J. Lengner, Yi Xing, and Russ P. Carstens. Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins. *Cell Reports*, 15(2):247–255, 2016.

[291] Janghoo Lim, Tong Hao, Chad Shaw, Akash J. Patel, Gábor Szabó, Jean-Fran¸cois Rual, C. Joseph Fisk, Ning Li, Alex Smolyar, David E. Hill, Albert-László Barabási, Marc Vidal, and Huda Y. Zoghbi. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, 125(4):801–814, 2006.

[292] Jinhu Yin, Yong Tae Kwon, Alexander Varshavsky, and Weidong Wang. RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes, interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule pathway. *Human Molecular Genetics*, 13(20):2421–2430, 2004.

[293] Shepherd H. Schurman, Mohammad Hedayati, ZhengMing Wang, Dharmendra K. Singh, Elzbieta Speina, Yongqing Zhang, Kevin Becker, Margaret Macris, Patrick Sung, David M. Wilson, Deborah L. Croteau, and Vilhelm A. Bohr. Direct and indirect roles of RECQL4 in modulating base excision repair capacity. *Human Molecular Genetics*, 18(18):3470–3483, 2009.

[294] Ya-Yu Chuang, Nhan L. Tran, Nicole Rusk, Mitsutoshi Nakada, Michael E. Berens, and Marc Symons. Role of synaptojanin 2 in glioma cell migration and invasion. *Cancer Research*, 64(22):8271–8275, 2004.

[295] R. Tacke, M. Tohyama, S. Ogawa, and J. L. Manley. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, 93(1):139–148, 1998.

[296] T. Hung, G. A. Pratt, B. Sundararaman, M. J. Townsend, C. Chaivorapol, T. Bhangale, R. R. Graham, W. Ortmann, L. A. Criswell, G. W. Yeo, and T. W. Behrens. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science*, 350(6259):455–459, 2015.

[297] Kazufumi Matsushita, Osamu Takeuchi, Daron M. Standley, Yutaro Kumagai, Tatsukata Kawagoe, Tohru Miyake, Takashi Satoh, Hiroki Kato, Tohru Tsujimura, Haruki Nakamura, and Shizuo Akira. Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *Nature*, 458(7242):1185–1190, 2009.

[298] Jian Liang, Jing Wang, Asim Azfer, Wenjun Song, Gail Tromp, Pappachan E. Kolattukudy, and Mingui Fu. A novel CCCH-zinc finger protein family regulates proinflammatory activation of macrophages. *The Journal of Biological Chemistry*, 283(10):6337–6346, 2008.

[299] Wenwen Fang, Yong Wei, Yibin Kang, and Laura F. Landweber. Detection of a common chimeric transcript between human chromosomes 7 and 16. *Biology Direct*, 7:49, 2012.

[300] Yasumasa Minoda, Kazuko Saeki, Daisuke Aki, Hiromi Takaki, Takahito Sanada, Keiko Koga, Takashi Kobayashi, Giichi Takaesu, and Akihiko Yoshimura. A novel Zinc finger protein, ZCCHC11, interacts with TIFA and modulates TLR signaling. *Biochemical and Biophysical Research Communications*, 344(3):1023–1030, 2006.

[301] Matthew R. Jones, Lee J. Quinton, Matthew T. Blahna, Joel R. Neilson, Suneng Fu, Alexander R. Ivanov, Dieter A. Wolf, and Joseph P. Mizgerd. Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nature Cell Biology*, 11(9):1157–1163, 2009.

[302] Haoran Wang, Mary X. Gao, Linda Li, Bin Wang, Naohiro Hori, and Kenzo Sato. Isolation, expression, and characterization of the human ZCRB1 gene mapped to 12q12. *Genomics*, 89(1):59–69, 2007.

[303] Alison Galloway, Alexander Saveliev, Sebastian Ł ukasiak, Daniel J. Hodson, Daniel Bolland, Kathryn Balmanno, Helena Ahlfors, Elisa Monzón-Casanova, Sara Ciullini Mannurita, Lewis S. Bell, Simon Andrews, Manuel D. Díaz-Muñoz, Simon J. Cook, Anne

Corcoran, and Martin Turner. RNA-binding proteins ZFP36l1 and ZFP36l2 promote cell quiescence. *Science*, 352(6284):453–459, 2016.

List of Figures

Danke!

First and foremost, I would like to thank my supervisor, Pål Westermark, for his guidance and his support, the inspiring conversations and the freedom he provided to follow my own way. Second, I would like to thank all the people who broadened my view with regards to science and, specifically, biology: Hanspeter Herzel, who often brought with his experience new aspects into the conversations, Rainer Machné, who taught me at least one quarter of my biology knowledge during long breakfast conversations, Carla Green, who gave me the oppurtunity to fulfill a scientific dream and study mitochondria, and last but not least Achim Kramer and his group, in whose journal club I learned to read Westernblots and other experimental data. I thank the Institute for Theoretical Biology to provide a nuturious, yet enjoyable environment for conducting and discussing research. A special mention goes to Andreas Hantschmann, Tiziano Zito, Rike-Benjamin Schuppner, Elvira Lauterbach and Karin Winkelhöfer for helping with computers, keys, contracts and so on.

I thank Pål, Hanspeter and Felix Naef for taking their time and evaluating this thesis, and Peter Hammerstein and Leonie Ringrose for being part of my commitee.

Apart from this scientific input I would like to deeply thank my family and my friends for their support and endless encouragement, especially Theresa, Kiri (thank you for reading all this!), Neta, Franzi and Grisha.

And finally, I thank Rainer for his endless patience, the critical comments on this thesis, the title (!) and last, but not least for his love.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 11.01.2017 Sarah Lück